Stochastic modelling and analysis of construction processes

Darren Graham

PhD Thesis – University of Edinburgh

June 2005

Abstract

Construction projects frequently overrun and finish over budget. This is in part due to the lack of control that construction practitioners – planners and managers – have over the construction schedule at a process level. Contemporary planning tools, such as critical path method (CPM) are adequate for planning at a project level, but are not of sufficient detail to help a practitioner to allocate resources to a process to maximise the performance. Thus the aim of this project was to develop a practical computer-based model to enable practitioners to plan projects at a process level and to improve their projects. This research has focused upon a specific type of construction process – those that are stochastic and cyclical, because such processes are: difficult to predict and hence control; and they are widespread throughout construction projects. Examples of processes that are stochastic and cyclical are crane operations, formwork erection and scaffold erection. Initially, a focus was placed on the process of ready-mixed concrete (RMC) supply. A significant amount of data was available for this process, collected through the observations of a previous PhD student.

This data was rigorously analysed to: determine if the process requires a non-linear stochastic solution, which it does; determine if the data was valid – a method of identifying outliers in a simple manner was devised to do this; determine the key variables for use in modelling. Guided by this statistical analysis, a discrete event simulation (DES) model of the RMC supply process was developed and validated. This model could not provide accurate estimates of the process, due to the need for a user define the probability distributions that represent the process. This is a complex requirement for a construction practitioner and the issue became known as the complexity problem and a solution was sought in the field of artificial intelligence. Case-based reasoning (CBR) provides solutions to new problems using knowledge of past ones – exactly the role of a practitioner in the above simulation model. Thus CBR was investigated as a solution to the complexity problem. CBR, when based upon a novel knowledge retrieval method, proved to be capable of solving the complexity problem.

A hybrid model, *CBRSim*, was developed to fulfil the original aim of this thesis. In this model, CBR is used to select the probability distributions to represent the process, given a set of user-defined operating conditions. These probability distributions are then supplied to a discrete-event simulation model that accurately recreates the process under the user-defined operating conditions. *CBRSim* was validated and found to be able to predict the productivity of the RMC supply process (volume delivered per hour) to within +/- 3% accuracy. Experimentation with various aspects of the process, such as resource allocation or supplier selection is performed to help illustrate the practical nature of the model.

CBRSim was then applied to another stochastic and cyclical process: earthmoving. *CBRSim* was found to be more accurate in modelling earthmoving than RMC supply, indicating a generic modelling capability.

The findings of this project have been widely disseminated through a variety of methods: 5 papers submitted to international, peer-reviewed journals (three of which have been published; one has been accepted) and 6 papers published at international conferences. Additionally, the work has been incorporated into and expanded via 12 undergraduate honours theses supervised by this author. It is likely that these students have been made aware that this problem exists and will hopefully utilise the methods proposed in this thesis in practice, to improve the construction industry.

Acknowledgements

There are many people to whom I am grateful. First I would like to thank Dr. Simon Smith for his unflappable belief in the project, his sound advice regarding the content and direction of the project, his encouragement for me to publish at an early stage, and for his door being ever open. Next, I would like to thank Dr. Paul Dunlop for his sterling work, which was a precursor to this project. When I was an undergraduate Paul raised my interest in this project and his work gave me a platform upon which to launch my own research. I would also like to thank all of the students whose projects I have co-supervised over the past three years: Jack, Maria, Richard, Julian, Adam, Paul, Harry, Doug, Pete and Neil. I learned a lot from the insightful comments and challenging questions of these students.

I would also like to thank the academics at the University of Edinburgh that I worked with over the course of this research. In particular: Dr. Martin Crapper for his sound advice on case-based reasoning and for permitting me to assist in the supervision of his students, Alison and Jamie; Prof. Bernie Mulgrew and Dr. Asif Usmani for giving me a tough first-year review – which made me think deeply about the purpose of the research, hopefully resulting in a better project.

I also wish to thank Prof. Iris Tommelein of the University of California, Berkeley for accepting me as a visiting scholar, for guiding my research with tremendous vision and allowing me to audit her excellent course on supply chain management and lean construction. Also, I thank the lean research group of UC Berkeley for permitting me to participate in their meetings. Final thanks to the EPM students for making me feel at home, showing me around the Bay area and working so hard – mimicking your work ethic helped drive me on to the end of this research project.

I am extremely grateful to Doug Forbes, construction manager at MJ Gleeson, for his contributions to the definition of model variables. This work helped make the models capable of being practically implemented.

I will be ever grateful to my family for their support and encouragement and those distracting squash tournaments and dinners.

The last word goes to Jen; you've made me laugh all the way through. Thanks, my love (∞!).

Table of Contents

1 THESIS OUTLINE AND RESEARCH CONTRIBUTIONS 1
1.1 Introduction 2
1.2 Thesis structure 2
1.3 Philosophy of this research project and the boundary of the work 4
1.4 Contributions of this thesis to the research community 9
1.5 Published work 11
2 IMPROVE CONSTRUCTION, IMPROVE SOCIETY 13
2.1 Introduction 14
2.2 Why does the UK construction industry need to improve? 14
2.3 Why is the industry under-productive? 17
2.4 Construction project hierarchy 18
2.5 What has been done to improve the UK construction industry? 19
2.5.1 Macroscopic level initiatives 19
2.5.2 Microscopic level initiatives 23
2.6 Problems to be addressed in this thesis. 24
2.7 Project aims 24
2.8 Project evolution 25
2.9 Chapter conclusions 26
3 READY-MIXED CONCRETE IN UK CONSTRUCTION PROJECTS 29
3.1 Introduction 30
3.2 How RMC is procured in major construction projects 30
3.3 The RMC supply and utilisation processes 31
3.3.1 Typical Preparation Work 32
3.3.2 The ready-mixed concrete supply process 34
3.3.3 The ready-mixed concrete utilisation process 34
3.3.4 Assumptions made in defining the above processes 35
3.4 Chapter conclusions 38
4 THE COLLECTION OF DATA 39
4.1 Introduction 40
4.2 Available methods of data collection 40
4.3 Method of collection of the original data sets 42
4.4 The sources and dimensions of the original data sets 47
4.5 Evolution of the project data 49
4.6 Chapter conclusions 51
5 STATISTICAL ANALYSIS OF THE PROCESS DATA 53
5.1 Introduction 54
5.2 The variables contained in the collected data 55
5.3 The underlying distributions contained in the data 56
5.3.1 A method for calculating the distribution of the data 57

5.3.2 Distribution analysis results 57
5.4 Dealing with problems contained in the data and reducing the variables 58
5.4.1 Dealing with missing values 59
5.4.2 Dealing with outliers in the data 60
5.4.3 Dealing with non-numerical variables 68
5.4.4 Dealing with correlated variables 69
5.5 Examining the nature of the relationship between the dependent and explanatory variables of the process 71
5.5.1 Identification of a systematic relationship between the variables 72
5.5.2 Quality of the linear fit 73
5.5.3 Confirming a non-linear relationship for the other dependent process variables 74
5.5.4 A further examination of the non-linear relationship between the process variables 75
5.6 Choosing model variables using a modelling framework 77
5.6.1 Modelling framework stages one and two: classifying the 'known' variables and listing them for further consideration 78
5.6.2 Modelling framework stage three: LS regression analysis of the remaining variables 79
5.7 Chapter summary and conclusions 80
6 SIMULATION MODEL DEVELOPMENT AND VALIDATION 81
6.1 Introduction 82
6.2 What is discrete-event simulation? 83
6.2.1 Components of a discrete-event simulation model 83
6.3 Discrete-event simulation use in construction 88
6.3.1 Simulation languages 90
6.3.2 General-purpose programming languages 93
6.3.3 A spreadsheet package 94
6.4 Simulation model development 94
6.4.1 Model verification 97
6.5 Fitting probability distributions to the six times of interest 98
6.5.1 Creating an input data histogram 99
6.5.2 Parameter estimation 100
6.5.3 Plot PDFs against input data histogram 101
6.5.4 Visual assessment of fit 101
6.5.5 Goodness-of-fit tests 102
6.5.6 Assessment of the fit of PDFs to the times of interest 103
6.6 Model validation 105
6.7 Chapter conclusions 107
7 A SOLUTION TO *THE COMPLEXITY PROBLEM* 109

7.1 Introduction 110
7.2 What is case-based reasoning (CBR) and why is it used in this project? 111
7.3 Case base design 114
7.4 Indexing mechanism 115
7.4.1 Crisp method *116*
7.4.2 Defining the crisp index sizes *118*
7.4.3 Fuzzy method *125*
7.4.4 Defining the fuzzy index sizes (or fuzzy membership functions) *127*
7.5 Retrieval mechanism 134
7.5.1 Variable weighting calculation *135*
7.5.2 Variable weighting sensitivity analysis *136*
7.5.3 Crisp retrieval mechanism *138*
7.5.4 Fuzzy retrieval mechanism *139*
7.5.5 Cluster-based retrieval mechanism *140*
7.6 CBR model development and validation 142
7.6.1 CBR model development *143*
7.6.2 CBR model validation *143*
7.7 Chapter Summary and Conclusions 144
8 THE HYBRID MODEL, *CBRSim* 147
8.1 Introduction 148
8.2 The hybrid model, *CBRSim* 149
8.2.1 CBR Module *149*
8.2.2 DES module *152*
8.3 Model Validation 153
8.3.1 CBRSim model validation *154*
8.3.2 Model cross-validation *156*
8.4 Limitations of *CBRSim* 159
8.5 An example to illustrate how to use *CBRSim* 160
8.6 Chapter conclusions 162
9 *CBRSim* AS A DECISION SUPPORT TOOL 163
9.1 Introduction 164
9.2 Decision 1: How many truck mixers are required? 164
9.2.1 Results from the truck mixer experiment *165*
9.2.2 A brief analysis of the experimental results *167*
9.2.3 Making the decision (on truck mixer quantity) given the project constraints *168*
9.3 Decision 2: RMC supplier selection 171
9.3.1 Project details *172*
9.3.2 Supplier details *173*
9.3.3 Simulating the realistic cost and delivery rate offered by each supplier. *174*

9.3.4 Simulation results *175*
9.3.5 Comparison of simulation results *178*
9.3.6 RMC supplier selection *180*
9.4 Chapter conclusions 180
10 ESTIMATION OF OTHER CYCLICAL, STOCHASTIC PROCESSES .. 183
10.1 Introduction 184
10.2 Modelling the earthmoving process 184
10.2.1 Defining the earthmoving process *185*
10.2.2 Data available for modelling *186*
10.2.3 Variable selection *186*
10.2.4 CBRSim model development *187*
10.2.5 CBRSim model validation *188*
10.3 Further cyclical, stochastic construction processes that could be modelled 190
10.3.1 Asphalt paving process *191*
10.3.2 Structural steel erection *191*
10.4 Chapter conclusions 192
11 CONCLUSIONS AND FUTURE WORK 193
11.1 Introduction 194
11.2 Summary of the original problem and problem evolution 194
11.3 Data availability and pre-modelling analysis 195
11.4 Model development, evolution and validation 199
11.5 Use of the model 205
11.6 Limitations of this research work and recommendations for future work 206
11.7 Learning outcomes from this project 208

REFERENCES 211

List of Figures

Figure 2.1: UK construction industry work value and productivity (after DTi, 2004). 17
Figure 2.2: Construction hierarchy (after Thomas et al., 2002) 19
Figure 3.1: Procedure of RMC procurement (Pullen et al., 2000) 31
Figure 3.2: The RMC supply and utilisation processes 33
Figure 3.3a: A concrete batching plant 35
Figure 3.3b: A truck mixer queuing on a project site 36
Figure 3.3c: A truck mixer discharging its load into a concrete pump 36
Figure 3.3d: Truck mixers queuing at a batching plant 37
Figure 3.3e: RMC being placed on a project site 37
Figure 3.3f: RMC being compacted on a project site 38
Figure 4.1: The RMC supply process 44
Figure 4.3: The locations of East of Scotland construction projects 48
(after, Collins Bartholomew Ltd., 2003) 48
Figure 4.4: The location of the motorway viaduct project 50
(after, Collins Bartholomew Ltd., 2003) 50
Figure 5.1: %age of times the distributions were ranked as most representative 58
Figure 5.2: The effect of an outlier on linear (LS) regression 62
Figure 5.3: The effect of clustering on normality of the process data 67
Figure 5.4: Scatter plot of grass min. temperature against daily min. temperature 71
Figure 5.5: LS regression estimation of productivity of the RMC supply process 73
Figure 5.6: Scatter plot of fit residuals against estimated truck-mixer utilisation 74
Figure 5.7: Scatter plot of fit residuals against estimated truck-mixer utilisation 75
Figure 5.8: Scatter plot of truck-mixer utilisation regression residuals vs. month of pour 76
Figure 5.9: Modelling framework that was followed in model development .. 77
Figure 6.1: Relationship between discrete-event simulation model components (after Law and Kelton, 2000) 86
Figure 6.2: An arrival event routine (after Law and Kelton, 2000) 87
Figure 6.3: A departure event routine (after Law and Kelton, 2000) 88
Figure 6.4: RMC supply process (with times of interest) 95
Figure 6.5: PDF vs. input data histogram (for position time) plot 101
Figure 6.6: Assessment of fit of PDFs to the travel to site time data 104
Figure 6.7: Correlated inspection results from the validation of *MatSim* 106

Figure 7.1: The parts of a CBR model and the interaction between them 113
Figure 7.2: An example view of the case base 115
Figure 7.3: Example crisp membership functions 117
Figure 7.4: Crisp indexing mechanism as defined by Sturge's rule 119
Figure 7.5: Crisp indexing mechanism as defined by the square-root rule 120
Figure 7.6a: Frequency graph of travel distance indexed into 3 indices..... 122
Figure 7.6b: Frequency graph of travel distance indexed into 6 indices 122
Figure 7.6c: Frequency graph of travel distance indexed into 9 indices..... 123
Figure 7.6d: Frequency graph of travel distance indexed into 12 indices 123
Figure 7.6e: Frequency graph of travel distance indexed into 15 indices..... 124
Figure 7.7: Crisp indexing mechanism defined by visual assessment..... 125
Figure 7.8: Fuzzy method of classifying ages in the example 126
Figure 7.9: Methodology of the iterative definition of fuzzy membership functions 129
Figure 7.10a: Chosen fuzzy membership functions for travel distance..... 130
Figure 7.10b: Chosen fuzzy membership functions for the number of truck mixers 130
Figure 7.10c: Chosen fuzzy membership functions for total pour volume..... 130
Figure 7.10d: Chosen fuzzy membership functions for maximum temperature 131
Figure 7.13: Comparison of productivities in weightings sensitivity analysis138
Figure 8.1: Methodology of *CBRSim* model 150
Figure 8.2: The activities within the RMC supply process 150
Figure 8.3: A sample of the case base from *CBRSim* 152
Figure 8.4: Graphical user interface of CBRSim 152
Figure 8.5: *CBRSim* estimated vs. observed productivity 154
Figure 8.6: CBRSim estimated vs. observed duration 155
Figure 8.7: Errors produced by each model in estimating productivity..... 157
Figure 8.8: Errors produced by each model in estimating duration 157
Figure 8.9: Errors produced by each model in estimating direct costs..... 158
Figure 8.10: Errors produced by each model in estimating truck mixer utilisation 158
Figure 8.11: Errors produced by each model in estimating on-site resources utilisation 159
Figure 9.1a: Productivity estimates from the truck mixer experiment 166
Figure 9.1b: Duration estimates from the truck mixer experiment 166
Figure 9.1c: Cost estimates from the truck mixer experiment..... 167
Figure 9.1d: Resource utilisation estimates from the truck mixer experiment167
Figure 9.2a: Duration estimates from the truck mixer experiment 169
Figure 9.2b: Cost estimates from the truck mixer experiment 170
Figure 9.2c: Resource utilisation estimates from the truck mixer experiment 171

Figure 9.3: The Transamerica Pyramid, San Francisco, California, U.S.A..... 173
Figure 9.4a: Productivity estimates for Company B (base pours) 175
Figure 9.4b: Duration estimates for Company B (base pours) 175
Figure 9.4c: Direct cost estimates for Company B (base pours) 176
Figure 9.4d: Resource utilisation estimates for Company B (base pours) 176
Figure 10.1: The earthmoving process 185
Figure 10.2: Variable selection made using the modelling framework (of chapter 5) 188
Figure 10.3: A sample from the case base of *CBRSim* 188
Figure 10.4: Correlated inspection results of *CBRSim* productivity estimates 189
Figure 10.5: Correlated inspection results of simulation productivity estimates (Smith, 1995) 189
Figure 10.6: Asphalt paving process 191
Figure 10.7: Structural steel erection process 192

List of Tables

Table 1.1: In-print or accepted publications resulting from this project 11
Table 2.1: The scope for sustained improvement (after Egan, 1998). 22
Table 5.1: The variables contained in the RMC supply process data set 56
Table 5.2: Missing values LS regression results 60
Table 5.3: Data points with large Penrose distances from the clusters in the data 67
Table 5.4: Non-numerical variables linear regression results 69
Table 5.5: Pearson correlation coefficients of the RMC supply process variables 70
Table 5.6: ANOVA table for linear regression analysis of the RMC supply process 72
Table 5.7: Steps one and two of the modelling framework applied to the process 78
Table 5.8: The variables to be used in model development 79
Table 6.1: Literature survey and review of 5 journals from 2002 to 2005 90
Table 7.1: Indices defined by visual assessment for travel distance (all in miles) 121
Table 7.2: Best number of indices identified through visual assessment 124
Table 7.3: Variable weightings calculated using linear regression 136
Table 7.4: Top five ranked cases in weightings sensitivity analysis 137
Table 7.5: Details of developed CBR models 142
Table 7.6: CBR model validation results 144
Table 8.1: Summary statistics of *CBRSim* validation (%age difference between model estimated and observed process performance measure) 155
Table 8.2: Input values of the variables of the CBRSim model 160
Table 8.3: The distributions (and parameters) suggested by CBR to solve the example 161
Table 8.4: Process performance estimates produced by CBRSim in the example 162
Table 9.1: Input values of the variables of the *CBRSim* model 165
Table 9.2: Results of the truck mixer experiment 165
Table 9.3: Tender bids submitted by the RMC suppliers (£ and m^3 in brackets) 173
Table 9.4: Assessment of estimates for Company B (base pours) 177
Table 9.5: Company B performance estimates 177
Table 9.6: Performance estimates for Companies A and C 178
Table 9.7: Total cost estimates by RMC supplier 179
Table 9.8: Total duration estimates by RMC supplier 179
Table 9.9: Truck mixer utilisation estimates (%) by RMC supplier 180

List of Appendices

APPENDIX A: MORE ABOUT LEAN CONSTRUCTION 225
APPENDIX B: APPLICATION OF THE SIMPLE OUTLIER IDENTIFICATION METHOD 235
APPENDIX C: THE RMC SUPPLY PROCESS VARIABLES 241
APPENDIX D: SOURCE CODE OF COMPUTER MODELS 247
APPENDIX E: CALCULATION OF THE PROCESS PERFORMANCE MEASURES 273
APPENDIX F: FIT OF PROBABILITY DISTRIBUTIONS TO THE TIMES OF INTEREST 277
APPENDIX G: VALIDATION OF *ExcelSim* 281
APPENDIX H: MATHEMATICAL NOTATION 283
APPENDIX I: PUBLICATIONS 285

CHAPTER 1

THESIS OUTLINE AND RESEARCH CONTRIBUTIONS

In order to introduce the work within this thesis, it is important to introduce the thesis structure itself. This chapter will firstly outline the thesis – it introduces each chapter in terms of the main problems the thesis addresses.

This thesis also intends to meet the needs of the research community and this chapter will indicate which contributions to the knowledge and understanding of the field of construction management have been made. This is demonstrated by highlighting the seven articles published to date.

1.1 Introduction

A fundamental aim of this thesis is to offer a new way for construction projects to be planned and controlled. Current critical path method-style project planning is not sufficiently detailed to aid a construction site manager to plan and control construction at a process level. This means that key activities (work-tasks) are not being performed with a maximum efficiency, at a significant cost, in terms of time and money, to resource suppliers, sub-contractors, general contractors and ultimately clients.

The method proposed in this thesis is to use simulation modelling as a means of permitting a site manager to easily compare alternative allocations of resources and determine the impact each one is likely to have on time and cost. This methodology is not new – it has been proposed in the past, but remains widely unaccepted in industry. What is new is that the tool that was developed and validated in this thesis has the potential to make simulation modelling an industry-wide standard for planning construction at a work-task level.

The main problem with the application of simulation modelling in construction is that the user needs to possess knowledge of the statistics underlying a process. Case-based reasoning was investigated in this research as a method of removing this need. It works by taking the underlying statistics from a past example of a process that is similar to the current situation and feeds them into a simulation model to allow experiments to be performed to determine how changes in resource allocation affect the outcome of a process. This hybrid case-based reasoning – simulation model is known as *CBRSim*, and it was developed and validated using data collected from real construction projects.

The practical application of *CBRSim* is demonstrated in two examples in this thesis, and it was shown that the model could help a practitioner make decisions about planning construction at the work-task level, and affect a reduction in the time and cost of, not only the operation being planned, but a whole construction project.

1.2 Thesis structure

The structure of this thesis is as follows:

- **Chapter two: improve construction, improve society**. This chapter contains a review of the state of the construction industry – and a need is identified for the industry to

increase its productivity to meet the challenges of today's construction market. Then the problem being addressed in this research and aims of this project are discussed. Subsequently, a description of how the project evolved is provided.

- **Chapter three: ready-mixed concrete in UK construction projects**. A description of how ready-mixed concrete (RMC) is procured, delivered and used in construction projects. The process of RMC supply was used as a test case for developing and validating the hybrid model discussed above.
- **Chapter four: the collection of data**. From an early stage in this project the importance of using real data from construction projects was evident. This chapter outlines the data collection efforts made in this project, including a description of the type, quantities and sources of data used in this project.
- **Chapter five: statistical analysis of the process data**. A statistical analysis of the project data was undertaken to ensure that the data was in a standard form and ready for use in modelling. For instance, the original data set contained missing values, non-numerical variables and outliers that needed to be dealt with prior to modelling. Also, standard statistical techniques were going to be used in the modelling process, for example to help select the key variables and these techniques often require data to be normally distributed – an examination of the distribution of the collected data was made to ensure that it was in fact normally distributed. Then an examination was made to determine whether a complex, non-linear solution was required to adequately model the process, as a more simple solution would be preferable. However, it was found that a non-linear, stochastic solution is necessary. Finally, having refined the data to a state where it was ready for use in modelling, a framework was developed to ensure that the variables to be used in modelling were practical – a construction practitioner would be able to know them at a specified date in a project. Under this framework the variables for use in modelling were defined.
- **Chapter six: simulation model development and validation.** A simulation model of the RMC supply process was developed and validated. This model was found to be too complex for a construction practitioner to use, as the accuracy of the estimates produced appear to be closely related to the quality of the statistical inputs that drive the simulation model. This issue became known as *the complexity problem* and a solution was sought in this research project.

- **Chapter seven: a solution to the complexity problem.** Case-based reasoning (CBR) was investigated as a potential solution to the problem discussed above regarding the need for a construction practitioner to possess knowledge of the underlying statistics of a process, given a set of operating conditions. CBR models are generally developed heuristically, with a number of different ways of storing and retrieving past information for use in new problems. This chapter explores many methods of performing these tasks, and then the best combination of these methods is identified in a validation process. The optimum indexing and retrieval method is subsequently used in the hybrid model, *CBRSim*.
- **Chapter eight: the hybrid model, *CBRSim*.** The hybrid model, *CBRSim* is introduced and validated. Furthermore, *CBRSim* is compared with the simulation model of chapter 6 and CBR models of chapter 7 for both accuracy and reliability of estimates. *CBRSim* was found to provide better estimates, in terms of both accuracy and reliability than the other two models.
- **Chapter nine: *CBRSim* as a decision support tool.** Two examples of how the model can aid a construction practitioner to make decisions about an operation are presented.
- **Chapter ten: estimation of other cyclical, stochastic processes.** The generic ability of *CBRSim* to model other cyclical, stochastic processes is explored by applying the model to the earthmoving process. Other processes that could be modelled by *CBRSim* are then outlined.
- **Chapter eleven: conclusions and future work.** Thesis conclusions and suggestions for future research.

1.3 Philosophy of this research project and the boundary of the work

In any project it is important to clarify what the boundary to that project is. A large factor, in determining what the boundary to a project is, is the philosophy being followed in carrying out the work of the project. This is because following one philosophy will naturally exclude, or place outside the project boundary, matters that are important to philosophies other than the one being followed. Thus this section sets out the philosophy followed in this research project and the consequent boundary of the work.

There are two dominant philosophies in construction management research: rationalist (also known as positivist) and interpretative. In the mid to late 1990s a debate was held in the construction management community, with the journal Construction Management and Economics acting as the 'sounding board' for the debate. The following is an analysis of this debate and then the philosophy followed in this work is stated and explained and a boundary to the project is drawn.

In 1995 Seymour and Rooke (1995) analysed the research being performed in the field of construction management and found that this research was largely dominated by a set of methodological assumptions and techniques which they called the 'rationalist paradigm'. Seymour and Rooke (1995) claim that 'chief among these assumptions and techniques is the attempt to import into management research the distinction which is drawn in science between subjective experience and objective reality'. They went on to assume that 'the use of this analytic device is intended to produce a neutral description of some field of human activity, with an aim to identify the constituent elements of a system, show how they interact and show how the system as a whole interacts'.

Seymour and Rooke admit that the attraction of this paradigm is that it seems capable of providing specific and scientifically valid statements about the probability of human effectiveness and it offers to show how key variables are related and how they can be managed to achieve successful outcomes.

However, Seymour and Rooke's contention with the above approach is that rationalist assumptions and ways of thinking about organisation have become institutionalised, deeply embedded within the culture of research and thus cannot produce the solutions to the problems the industry currently faces. Seymour and Rooke also felt that a rationalist approach is the cause of too much 'bad' research, propagated by researchers 'hiding' behind the evidence produced by statistics.

Seymour and Rooke (1995) suggest an 'interpretativist approach' to research is more favourable. Where, an interpretative approach involves a researcher addressing a problem by not personally defining any boundaries around the problem or solution (Seymour and Rooke, 1995), unlike in the rationalist approach where the researcher hypothesises a solution to a

defined problem. Instead in the interpretative approach the boundaries around a problem and solution are defined by the people involved in the problem themselves. In the field of construction management these people would be the workers in the project, the site engineers, engineering designers, architects and the many others who are involved in a project. The implementation of the 'interpretativist approach' is akin to brain storming (gathering ideas without any analysis of how good the ideas are thought to be), with the people involved, on the problems and solutions that are to be addressed in a research project. Seymour and Rooke (1995) felt that this approach would result in gaining an understanding of a situation or problem without any 'bias' being added by the researcher carrying out the work.

The paper in 1995 did not 'light many fires' in the research community. Rather it took a note in an early 1997 edition of the journal Construction Management and Economics to spark debate. This note (Seymour et al., 1997) was a short version of the paper published in 1995, but it carried many of the same messages and was entitled a call for debate.

Seymour and Rooke's wishes were certainly granted, with no less than six replies being published in the same journal over the period 1997-1998. The first opponent of Seymour et al. was Runeson (1997) who criticised the arguments of Seymour et al. saying that mainstream methodology (the rationalist approach) has been modified to deal with the intricacies of construction management research and these methods have been instrumental in substantial advances in science. Runeson goes further, stating that Seymour et al is using an argument that is 'ostensibly anti-scientific', and although it has been around for a long time, there are no positive achievements to suggest that we would benefit from adopting it. Runeson closes by stating that positivist methods are the best insurance against bad research.

The views of Runeson are supported by numerous authors such as Raferty et al. (1997), Harriss (1998), Wing et al. (1998) and Raferty (1998). These authors argue for a scientific, positivist solution, generally concluding that an interpretative approach to research can help to introduce alternative hypotheses which can then be tested using a rationalist approach. The 'interpretativist' lobby (Seymour et al.) replied to the above critiques of their arguments in Rooke et al. (1997) and Seymour et al. (1998).

The author of this thesis remains unconvinced by the interpretative approach to research and the arguments proposed by Seymour et al. The reasons for this are three-fold:

1. A key assumption of the argument of Seymour et al. (1995) is that causal relationships do not exist in construction projects and hence management, nor indeed do they exist in any management, because these activities involve humans and they act independently. This is clearly different to the founding assumptions of science that there is a causal relationship in natural systems.

   The work presented in this thesis involved the study of stochastic construction processes, which involve the coming together, or interaction, of humans, machines, materials and nature.

   It is difficult to agree with Seymour et al. that there can be no causal link between the variables of the process, just because there is human interaction involved. Rather, it is likely that a stochastic process is performed in a particular manner due to the causal links between the human, machine, materials, and natural systems. For instance, if there is a high wind speed, then it is not possible to operate a concreting pump safely. This is an interaction between natural and machine systems. It could be argued that this interaction has no impact on the stochastic process of concreting, because a human could continue regardless of safety. This is true, but only to an extent. The concrete pump would prove to be much less user-friendly under severe weather conditions than in a temperate situation, thus the natural system and machine system will impact on the process regardless of the interaction with the human. The author firmly believes in seeking a causal link between the variables of the process.
2. A strong causal link between the variables of many stochastic processes has been reported in the recent past. An example of this work is provided by Dunlop (2005) in modelling the concreting process using linear regression analysis. This is work to be built upon, not rubbished because it utilises established methods of analysis and research.
3. Finally, and this may appear at first glance callous, the rationalist approach to research is preferable to the author of this thesis because it produces results. As reported later in this thesis, there is a problem in the relationship between the construction industry and construction researchers; there is a significant amount of

research performed and reported, but an insignificant amount of the proposals from researchers is adopted for use in practice by industry. To adopt an approach to research that has been, over time, shown to not produce many results (Runeson, 1997) could widen the gap between industry and the research community further. Seymour et al. (1995, 1997) assume that the research community are suggesting the wrong methodologies to industry, but it is possible that the industry is just unwilling to listen to the research community. Thus, reducing the volume of work would only act to weaken the voice of the research community, and subsequently reduce the adoption of its suggestions in industry. This argument is assuming that there would be no substantial increase in the quality of research produced through using an interpretative approach over a rationalist one, which has also shown to be true over the course of time (Runeson, 1997).

As stated above, the author of this thesis believes in adopting a rationalist approach to research. This approach involves stating a scientific hypothesis and trying to prove or disprove its correctness. The hypothesis proposed in this thesis is that probability theory can be applied to understand the relationships between the variables of stochastic, cyclical construction processes.

In adopting a rationalist approach in this research there is a tendency to place more weight upon quantitative methods. Thus, to place a boundary around this thesis, it would be fair to say that qualitative methods of analysis, such as systems maps, are outside this boundary, whereas quantitative methods of analysis are inside. Although qualitative methods are not considered in this work, this is not to say that qualitative data is ignored in the work presented in this thesis, but where it has been included it is often transformed into a numeric form for use with the quantitative methods used in the research.

To summarise, the boundary can be defined by stating what was included in this research: quantitative methods of analysis and data are used in seeking to prove the above hypothesis; qualitative methods of analysis are outside of the remit of this thesis; the use of qualitative data falls on the boundary, as this type of data is used, but only after transformation to numeric form.

1.4 Contributions of this thesis to the research community

The contributions of this research to the fields of construction management, stochastic modelling and artificial intelligence made by this thesis are deliberately extensive. These contributions, their locations in this thesis and, where applicable, their place of publication, are summarised as follows:

- An extensive literature review was undertaken and is presented **(chapters 2, 4, 5, 6, 7 and 8).**
- Working in collaboration with Prof. Iris Tommelein at University of California, Berkeley during a five-month scholarship, the supply chain of ready-mixed concrete was defined and then re-engineered in accordance with lean construction principles. This research work was expounded in three papers, one of which is under review with the journal *Construction Management and Economics*; the other two have accepted for publication at the *ARCOM (Association of researchers in construction management)* 2005 conference. These papers can be found in appendix I.
- A simple method of identifying outliers in data sets was devised. This may prove to be more practical than the commonly used method based upon the Mahalanobis distance measurement **(chapter 5).**
- A framework to ensure the production of practical models was developed **(chapter 5).** This was the subject of a paper presented at *ARCOM 2004*, which can be read in full in appendix I.
- The underlying statistics of the concreting and earthmoving processes were examined **(chapter 6)**. This investigation found that no individual distribution could represent the process in every instance. However, the lognormal distribution was found to provide the best representation on more occasions that any other – this is a new finding in a field that previously accepted the beta distribution as an adequate representative. This was the subject of a paper published in *Journal of Construction Engineering and Management, ASCE* in February 2005, which can be read in full in appendix I.
- Simulation was confirmed as being overly complex for use in the construction industry **(chapter 6).**

- The ability for CBR to adequately estimate the productivity of construction processes was explored **(chapter 7).** This was the subject of a paper published in *Advanced Engineering Informatics* in November 2004, which can be read in full in appendix I.
- This research identified that standard methods of forming histograms could prove useful in indexing cases in CBR modelling. This method was researched and found to be useful **(chapter 7).**
- The application of fuzzy indexing and retrieval methods in CBR models of construction was examined **(chapter 7).**
- The use of the decision tree-forming, ID3 algorithm as a retrieval method in CBR models was explored **(chapter 7).** This was the subject of a paper published in *Advanced Engineering Informatics* in November 2004, which can be read in full in appendix I.
- A novel, cluster-based retrieval mechanism for CBR models was developed. This is more efficient than standard mechanisms as it rejects the most unlikely cases from consideration and then focuses on choosing a solution from a greatly reduced group of cases **(chapter 7).**
- The product of this research has been to make simulation a more accessible tool for the construction industry. To this end, a hybrid model, CBRSim was developed that removes the need for a construction practitioner to hold knowledge about the statistics underlying a construction process to use simulation effectively **(chapter 8).** This was the subject of a paper published in *Civil Engineering and Environmental Systems* in June 2004 and presented at *ARCOM* 2003 (see appendix I).
- CBR was shown in this thesis as a successful method to remove the need for a user to possess knowledge of the statistics that underlie a process or system, thereby solving *the complexity problem.*
- In this research it was demonstrated how simulation may be used to aid decision making in the process of planning and controlling construction operations **(chapter 9).** This was the subject of a paper presented at *Construction Research Congress, ASCE* in April 2005, which can be read in full in appendix I.
- Additional publications have stemmed from this research, but not been included in the text. These include investigations into the application of neural networks to modelling construction processes, which were published at *ARCOM* 2004 and currently under review with *Automation in Construction* (see appendix I).

1.5 Published work

A total of 5 journal papers have been produced from the work in this PhD research project; to-date 3 of these papers has been published, one has been accepted for publication. In addition, 6 conference papers have been written from the work performed in this research project; 4 have been published in conference proceedings; 2 have been accepted for publication. Table 1.1 provides a summary of the publications that have stemmed from this research project that are either in-print or have been accepted. Full details of these publications are provided in appendix I.

Paper Title	Publication Title	Journal or Conference
Improving concrete placement simulation with a case-based reasoning input	Civil Engineering and Environmental Systems	Journal
Estimating the productivity of cyclic construction operations using case-based reasoning	Advanced Engineering Informatics	Journal
Lognormal distribution provides an optimal representation of the concrete delivery and placement process	Journal of Construction Engineering and Management, ASCE	Journal
Modeling the ready mixed concrete delivery system with Neural Networks	Automation in Construction	Journal
A hybrid model to improve the estimation of construction operations	19th ARCOM annual conference, University of Brighton	Conference
A method for effectively implementing construction process productivity estimation models	20th ARCOM annual conference, Heriot-Watt University	Conference
Concrete placing productivity using a novel neural network design	20th ARCOM annual conference, Heriot-Watt University	Conference
A model to facilitate the optimization of the concrete supply process	Construction Research Congress, ASCE	Conference
Cost of concrete placement for contractors	21st ARCOM annual conference, Salford University	Conference
Cost reduction through vertical integration of the in-situ concrete supply chain	21st ARCOM annual conference, Salford University	Conference

Table 1.1: In-print or accepted publications resulting from this project

CHAPTER 2

IMPROVE CONSTRUCTION, IMPROVE SOCIETY

The UK construction industry does not perform as adequately as it could. This chapter considers why this should be and how this impinges on the needs of society. It presents ideas as to where the problem originates and past initiatives that have been carried out to investigate it. This allows the aims of the research project and the questions that this thesis will pose to be considered – and the manner in which they can potentially be answered.

2.1 Introduction

Construction is important to everyone. The products of the construction industry are the structures in which we live, become educated, receive medical treatment and work, and the infrastructure that gives us an ability to travel between these places. A primary objective of this thesis is to outline ways to improve the performance of construction projects, with the hope that this will lead to better housing, schools, hospitals and public transport. In identifying the need to improve construction, focus shall be placed on the UK construction industry, but it is hoped that any methods of improving construction proposed in this thesis will prove useful on an international scale. Hence the title of this chapter reflects the optimistic underlying thought of this thesis.

Chapter two is structured as follows:

- **Section Two.** Some reasons why the UK construction industry should seek to improve upon its performance.
- **Section Three.** A theory of why the industry is under-performing.
- **Section Four.** A description of a typical construction project hierarchy.
- **Section Five.** An examination of what has been done or proposed to improve the performance of the UK construction industry. The main methods proposed have been at a project or organisational (macroscopic) level and at a (microscopic) process level.
- **Section Six.** Problems to be addressed in this project.
- **Section Seven.** The aims of this project.
- **Section Eight.** PhD project evolution.
- **Section Nine.** Chapter conclusions.

2.2 Why does the UK construction industry need to improve?

A significant amount of the work currently being undertaken by the UK construction industry is to, develop new and refurbish existing, Public services. This work is being financed by Public funds directly, Public-Private Partnerships (PPP) or through the Private Finance Initiative (PFI). Regardless of which method of finance is being considered, the end-result should be the same – the construction or rehabilitation of infrastructure (e.g. buildings, roads), in a timely, cost-effective and quality-driven manner, that permits the Public sector to offer excellent service to the people of UK. In summary, there is an onus on the UK construction industry to deliver "value" (time, cost and quality) to the people of the UK.

An example of this Public sector construction is provided by the Deputy Prime Minister, John Prescott's plan to deliver 200,000 new homes for "front-line" public sector workers and people on low incomes, in the Thames Gateway, in the south east of England, by 2010 (Weaver, 2002). The aim of building this social housing is to allow those with low household incomes the opportunity to own homes in areas where current house prices prevent it, but 200,000 new homes may prove to be a conservative estimate of the actual amount of homes required to achieve this aim. Research conducted by economists at Cambridge Econometrics estimated that more than 1 million new homes would be required, stimulating images of a booming construction industry (Denny and Halls, 2002). The view that John Prescott is underestimating the amount of new homes required in the south east of England has been supported by the Barker review to the UK Government on housing supply requirements (Barker, 2004).

Another example of Public sector construction work is found in the management of the UK's railways. In 2003 the UK Government and the private sector each spent £3.8 billion on the railways and the rail industry (DfT, 2004), on the following projects:

- Building a major new line (Channel Tunnel Rail Link).
- Upgrading the West Coast Main Line.
- Installing a train protection and warning system to improve safety.
- Modernising power supply south of the River Thames.
- Renewed 800 miles of track.

Network Rail – a non-profit organisation- was formed on 2nd October 2002 and its role was to perform the above work and to hold responsibility for the national railway infrastructure (Clark, 2002). The chairman of Network Rail gave an indication of the amount of work he thought remains to be completed on Britain's railways, when he called for an increase in investment from £15 billion to £22.7 billion over 5 years from 2003 (Clark, 2003).

This sustained building and refurbishment program should be the cause of satisfaction for the Government and the construction industry. After all, the Government will be giving the electorate what they desire - better public services (social housing and railways in the above examples) - resulting in popularity and the possibility of re-election; and the construction industry is getting a lot of work, resulting in more secure jobs and higher profits. The truth being, that neither of these entities are completely satisfied with the situation. This is because

the UK construction industry in its current mould is ill equipped to deal with the increasing workload created by this investment.

The consequence of this is an increase in competition from contractors from outside of the UK, especially as there is open competition within the European Economic Area. This increased competition could lead to the disappearance of UK contractors and less money being retained in the UK economy, which is significant to the UK Government; the construction industry is worth around 5.4% of the nation's Gross Democratic Product (GDP) (DTi, 2003).

Some evidence of the inability of the UK construction industry to cope with contemporary demand is shown in figure 2.1, which displays the performance of the construction industry, in terms of value of work and productivity, over the past 25 years (DTi, 2004) – where, productivity is a measure of the rate of work of the industry, which is represented by the monetary worth (output) of the work of each employee and value of work is indicative of the quantity of work being undertaken. Figure 2.1 shows that since 1993 the value of work that was completed every year by the industry grew on average by 1.3% per annum. During this same period the productivity of the construction industry grew on average by only 0.4% per annum. The above trends signify that the workload of the industry is growing, but the industry cannot work quickly enough to complete the workload. Focussing on the previous four years (2000-2004) is even more alarming, as the construction work value grew annually by 1.5% while productivity increased by only 0.375%. This expansion in work value is being fuelled by the UK Government's investment in the public sector (Construction Monitor, 2003) and therefore the stagnation of the construction industry's productivity adversely affects the delivery of these services to the electorate; the result being a dissatisfied Government and dissatisfied citizens.

Clearly, there is a problem with the productivity of the UK construction industry and it is a problem in which the United Kingdom's construction industry, Government and citizens have a vested interest. The questions posed next are why is this problem occurring and what is being done or proposed to solve this problem?

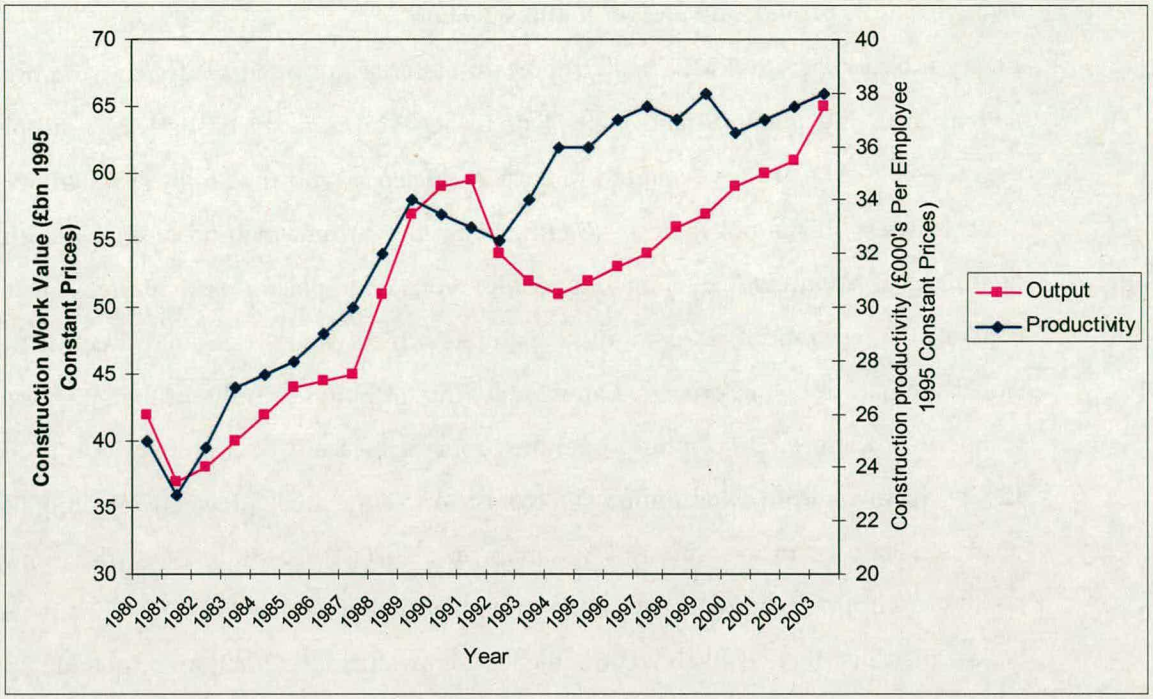


Figure 2.1: UK construction industry work value and productivity (after DTi, 2004).

2.3 Why is the industry under-productive?

The list of reasons why the UK construction industry is under-productive is extensive, and includes inefficient work and management practices, poor training methods and facilities, an inexperienced workforce and management, and an inability to attract and retain suitable workers.

Another cause of under-production - one that is considered to be extremely significant - is the propagation since the 1970s of construction management techniques that provide planning and control at the project level, such as the critical path method (CPM) or the line of balance (LOB). While these techniques can appear to be effective at an overall project level, they allow no provision for controlling (ordering, allocating, etc.) the resources required to perform individual operations (instances) of construction processes; they are simply not detailed enough (Halpin and Woodhead, 1976). So how are resources allocated to an operation? Currently, a construction practitioner makes an intuitive estimate of the likely productivity, duration and cost of the process given the operating conditions and a particular allocation of resources. Is this current method effective?

The following example provides an answer to this question:

> Ready mixed concrete (RMC) is an important material in contemporary construction, and as such the planning and control of the cyclical and stochastic[1] RMC supply process should be effective enough to ensure efficiency (and thus high productivity). Unfortunately this is not the case (Smith, 1998): the current method of choosing the quantities of resources (e.g. plant and people) required to place a particular volume of concrete (an operation) is an intuitive estimate which cannot take into account the stochastic nature of the process. On average, this technique results in the site-based resources (concrete pump, pump operator, concrete placers, etc.) remaining idle for 12% of the overall process duration at a cost of 14% of the total, adversely affecting the productivity of a process by 16% (Graham et al., 2004). A study of earthmoving, another cyclical, stochastic construction process revealed similar findings. Thus it can be assumed that there is likely to be some loss of productivity due to poor planning of construction processes. With cyclical, stochastic construction processes, like RMC supply and utilisation and earthmoving constituting up to 80% of construction projects, a small loss in productivity (of say 16%) in each of these processes can have a profound impact on the productivity of a whole project.

There is a problem with contemporary project planning at a process level, but where in a project's hierarchy is this level located? The next section provides an answer to this question.

2.4 Construction project hierarchy

To help clarify where processes fit into a construction project, a typical project hierarchy has been drawn up and is shown in figure 2.2. The hierarchy is in a top-down form with the project (e.g. construct a new hospital) at the top. The next level comprises the different disciplines involved in a project (e.g. civil engineering). Below this are the processes that comprise all of the work that must be completed by a particular discipline (e.g. earthmoving). The lowest tier contains the activities that comprise each of the construction processes (e.g. batch a concrete truck mixer).

---

[1] A stochastic process is characterised by a sequence of random elements (Oxford English Dictionary, 1996). A simple example of one of these 'random elements' is that the road conditions encountered by truck mixers will never be identical, resulting in variable journey times between the batching plant and the project site.

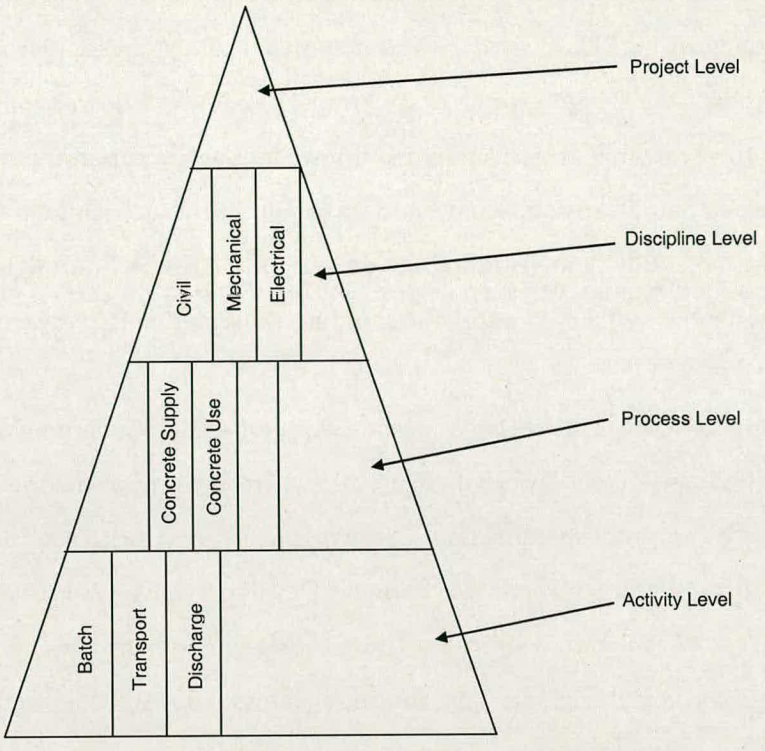


Figure 2.2: Construction hierarchy (after Thomas et al., 2002).

2.5 What has been done to improve the UK construction industry?

There has probably always been some concern for the UK construction industry's health, because it is so important to the national economy. This section is a summary of the key initiatives that have been promoted to improve the performance of the construction industry, by academic research, the UK Government and the industry itself. They were mainly applied with two views: macroscopic - initiatives applied at the industry, organisational or project level; or microscopic - initiatives applied at the process level.

2.5.1 Macroscopic level initiatives

The academic research community have identified the need for improvement in general construction practice and performance. In 1984, UK construction management researchers formed an association (ARCOM) to hold an annual conference to discuss the relevant issues of the day, and to promote innovative methods of improving the performance of the construction industry. A major issue faced by academic researchers in construction management is that little attention is paid by the industry to their suggestions, and only a small amount of research findings filter through to the construction industry. This is evident in the fact that

construction firms have reduced their research and development budgets (including funding paid to academics) by 81% over the 17-year period leading up to 1998 (Egan, 1998). This notwithstanding, the Engineering and Physical Sciences Research Council (EPSRC) purportedly fund research projects more readily when they are supported by industry – in the form of time and data (Dawood, 2005) – and this should drive academics to form partnerships with industry, hopefully demonstrating the benefits of the research through this collaboration and encouraging the industry to adopt the solutions presented in the research.

The following initiatives have been proposed by the UK Government to improve the construction industry. Over the past fifty years, there have been a series of major studies concerned with the problems and issues of worldwide construction industries (e.g. Simon, 1944; International Bank for Reconstruction and Development, 1973; World Bank, 1984). One of the earliest studies undertaken in the United Kingdom was by Sir Ernest Simon on 'The placing and management of building contracts' (Simon, 1944). This remit was broadened largely in 1962 when the Emerson Report was published entitled 'A survey of problems before the construction industries' (Emerson, 1962). This report was closely followed by Banwell (1964). These UK-based studies appear to be undertaken in cycles, which as they are commissioned by the Government of the day are most likely to be governed by political pressure.

Given the lack of major studies of the whole construction industry since 1964 and that the cycle-time of these studies appears to be roughly 20-30 years, the UK Government of 1993 announced a joint industry/Government commission to identify the flaws in construction practice and solutions to those flaws. This commission, led by Sir Michael Latham was to spend one year and up to £250,000 on the investigation and to report back to the UK Government, providing recommendations as to how construction practice may be improved upon. In 1994 Latham delivered his commission's report 'Constructing the Team', which was frank and scathing - it found the industry to be 'ineffective, adversarial, fragmented, incapable of delivering for its customers, and lacking respect for its employees' – it held in essence a similar sentiment, with the exception of the issue of lack of respect for employees, to reports by Emerson and Banwell. Latham realised this was the case and that fundamental change in the industry was required. In a bid to avoid another study whose findings are ignored Latham highlighted the problem by stating 'if this opportunity to improve construction is not taken, it

may not re-occur for many years, and a new report may be commissioned in the year 2024 to go over the same ground again.'

'Constructing the Team' was not just a wake-up call for the industry, it set an agenda and targets for reform. A number of initiatives were formed as a result of 'Constructing the Team': 'the establishment of the Construction Industry Board to oversee reform; 'The Construction Best Practice Programme' to encourage excellence in construction; 'The Movement for Innovation'; 'Constructing Excellence', a recently formed industry body to deliver the accelerating change agenda and to supply the construction industry with a support network to become productive and profitable (Constructing Excellence, 2004). Another report was commissioned, led by Sir John Egan in 1998.

The newly elected UK Government announced in 1997 that a new study of the construction industry would be conducted by a team to build upon the knowledge gained from the 'Constructing the Team' report of 1994. This study, called officially 'Rethinking Construction' or unofficially 'the Egan report' found that the UK construction industry when at its best, was excellent, with a capability to deliver the most difficult and innovative projects to match that of any other construction industry in the World (Egan, 1998). However, it also concluded that the industry was under-achieving: by having a low profitability; by investing too little in capital, research and development and training; and, by producing too many dissatisfied clients. 'Rethinking Construction' set seven targets for the industry (see table 2.1), including: reduce capital cost and construction time by 10% annually; and increase productivity by 10% per year. To achieve the seven targets, five 'key drivers of change' were proposed (Egan, 1998):

- Committed leadership
- A focus on the customer
- Integrated processes and teams
- A quality driven agenda
- Commitment to people

Indicator	Improvement per year
Capital Cost (All cost excluding land and finance)	Reduce by 10%
Construction Time (Time from client approval to practical completion)	Reduce by 10%
Productivity (Increase in value added per head)	Increase by 10%
Predictability (Number of projects completed on time and in budget)	Increase by 20%
Turnover and Profits (Turnover and profits of construction firms)	Increase by 10%
Defects (Reduction in number of defects on handover)	Reduce by 20%
Accidents (Reduction in number of reportable accidents)	Reduce by 20%

Table 2.1: The scope for sustained improvement (after Egan, 1998).

The Latham and Egan reports produced a whole raft of measures to improve the performance of the construction industry. They announced to the industry that clients must be made to feel satisfied with the product they are receiving in terms of the construction duration, cost and quality. It could be said, although neither report explicitly did so, that the proposals made by Latham (1994) and Egan (1998) are embracing the principles of lean production management. This is evident because lean production management: is stimulated by the same five 'key drivers of change' as identified by Egan (1998); and aims to produce satisfied clients through improved work practices. This fact was not acknowledged by the Egan Report (1998), which stipulated that the adoption of the methods used in lean manufacturing should be made, but only as one of a 'raft' of measures. In reality, Latham (1994) and Egan (1998), whilst using other terms and phrases, were advocating the use of lean principles to solve the problems of

the construction industry (Green, 1999). These principles when applied to construction are known collectively as lean construction.

Has lean construction been considered by any studies in the post-Egan era? The answer is yes. There have been several reports commissioned by a variety of institutions and most have at least considered lean construction as a method of improving future construction performance. In 1998, a report by CIRIA proposed that contractors should be selected by value, not minimum cost (Jackson-Robbins, 1998) – which is a fundamental principle of lean. Subsequently, another report by CIRIA entitled 'The handbook of supply chain management' suggested that lean methods should be adopted such as selecting suppliers by value, simplifying the supply chain and implementing continuous improvement methods (Holti et al., 2000). Finally, a report 'Modernising construction' was published by National Audit Office in 2001, which promoted the use of the lean methods stated in 'The handbook of supply chain management'.

Clearly lean construction is increasingly thought to be a fundamental element of the future success of the UK construction industry. What lean construction is and how the UK construction industry has reacted to it are discussed in detail in appendix A.

2.5.2 Microscopic level initiatives

It is thought in some quarters that improving the planning and thus performance of construction processes (using a microscopic level initiative) is a method of improving the overall performance of construction projects and the industry. These thoughts are not new, and have been researched for over three decades by academics internationally such as: Halpin (e.g. 1977; 1993) – used simulation to improve general construction processes; Carmichael (e.g. 1986) – used queuing theory to model the earthmoving process; Wales and Abourizk (e.g. 1996) – modelled the effect of weather on construction; Wakefield (e.g. 1998) – developed Petri net models of construction processes; and in the UK, Smith (e.g. 1999) – performed experiments on the earthmoving process using linear regression techniques. Such studies are reviewed in detail in chapters 6 and 7. In brief, most of these studies have focussed on modelling construction processes to a degree of detail to allow the impact of resource allocation to be assessed.

Unfortunately a fundamental drawback of the above line of investigation to date has been the lack of interest shown by industry in adopting the tools and techniques proposed by academia. This lack of interest appears to have been caused or at least propagated by: contractors not being aware of the problem; if they are aware of the problem then not being aware of the proposed solutions; if they are aware of the solutions then they are being dissuaded by the apparent complexity of the solutions being presented by academia; and finally, the solutions presented by academia may have been dismissed by industrialists because they place doubt upon their own judgement – a reliance must be made upon computers rather than humans.

2.6 Problems to be addressed in this thesis.

Given the discussion above on the under-performing UK construction industry and the implications of this for construction firms and wider society, the following problems shall be addressed in this thesis:

- Contemporary methods of construction project management provide insufficient control of the details of a project, below the project level. This is particularly evident when there is a need to allocate resources for use in operations of construction processes, and the only widely available method is to make an intuitive estimate of how many resources are required and how these resources will perform.
- It was determined in this study that poor resource allocation can result in a loss of productivity of up to 16% in cyclical, stochastic construction processes, which in turn can comprise up to 80% of a total project.
- Practitioners in industry have not readily accepted previous solutions proposed by academia to solve the above problem. They are often thought of as overly complex and difficult to use.

2.7 Project aims

Taking into account the above issues, the aims of this project can be defined as:

1. **To investigate methods to improve the performance of construction projects and the UK construction industry**. This could be achieved by developing a model that not only accurately estimates process performance measures (productivity, cost, etc.), but allows experiments to be undertaken which determine how process performance varies with the

resources allocated for use in an operation. This would alleviate the need for intuitive estimates to be made.

Currently, the construction industry is losing many practitioners that are highly experienced at estimating the performances of construction processes. The long-term benefit of using a model to provide such estimates is that a construction contractor can avoid the loss of knowledge incurred when an experienced estimator retires.

2. **To develop models capable of practical application.** The implementation of previous models of cyclical, stochastic construction processes have been hampered in part by the models being overly complex at the user end. Thus, any models developed in this PhD project shall attempt to avoid this problem and produce a practical, easy to use model.

3. **To learn more about the processes.** An accurate model that is capable of performing experiments can provide information about the processes being modelled. In this project the developed model can be used as a decision support tool. By using the model in this manner a great deal can be learned about stochastic, cyclical processes without the expense involved with performing a real experiment.

2.8 Project evolution

To begin, one cyclical, stochastic construction process shall be focussed upon in this PhD project, the supply of ready-mixed concrete (RMC). This is because:

- It is a classic example of client (truck mixer)/server (pump) system, lending itself to commonly used modelling methods such as simulation.
- Also, the process is not being planned and controlled adequately using contemporary methods.
- Finally, the majority of the data available is from observations of the RMC supply process.

The initial proposal for the form of the computer model is a discrete-event simulation (DES) model, based upon real-life data obtained from the observation of the RMC supply process. DES was considered because it has been used in the past to successfully model cyclical,

stochastic construction processes, and its ability to recreate the RMC supply process has been provisionally proven in past research by Smith (e.g. 1998).

The model evolved over the lifetime of this PhD project, with the realisation that a construction practitioner could not easily determine the statistics that represent the durations of the six activities that must be undertaken to perform a cycle of the RMC supply process. These durations are known in this thesis as the *times of interest* of the process. The statistics that represent these *times of interest* are presented in the form of probability distributions, and these distributions give a DES model the ability to re-create the rate at which concrete was supplied in the past. In summary, the accuracy of the above distributions is essential to the ability of a discrete-event simulation model to accurately recreate the RMC supply process. The requirement for a construction practitioner to know how the above probability distributions changed under different operating conditions was deemed to make the DES model too complex. This problem with DES is known in this thesis as *the complexity problem.*

Case-based reasoning (CBR), an artificial intelligence paradigm, has been used as a method to solve *the complexity problem.* CBR is used here to suggest the appropriate probability distributions to represent the *times of interest* of the RMC supply process for a given set of operating conditions – removing the need for a construction practitioner to do this; thus making the model more accessible and practical to construction practitioners. The final form of the RMC supply model is a hybrid between CBR and DES, with CBR providing an input and DES an output.

The application of this model is then extended to the earthmoving process. It is demonstrated that the model has the capability of being generic, thus expanding the potential impact that such a model can have upon the performance of construction and the wider community at large.

2.9 Chapter conclusions

The following conclusions may be drawn from the discussion contained in chapter 2:

- Public sector, PPP and PFI investment is driving the demand for construction services and the UK construction industry appears to be having difficulty in meeting this demand due to poor productivity levels. A solution to this problem needs to be found.

- The poor productivity levels appear to be propagated by contemporary construction management tools such as critical path method (CPM) that provide control at a project level. Such tools ignore the need for a construction practitioner to allocate resources to individual operations of construction processes, leaving this important decision to the intuition of the construction practitioner. This intuitive allocation of resources was found in this study to reduce productivity by 16%, in cyclical and stochastic processes (which can make up 80% of a project).
- This project will attempt to devise methods of improving the productivity of the construction industry by developing a model that aids the effective allocation of operational resources in cyclical, stochastic construction processes. Hopefully this will assist construction practitioners to make improvements in process, project and ultimately industry productivity.

CHAPTER 3

READY-MIXED CONCRETE IN UK CONSTRUCTION PROJECTS

To address the aims of the project, as discussed in chapter 2, a particular construction process has been used to develop and test models and theories. This process is *Ready-Mixed Concreting (RMC)* and this chapter introduces the process, its attributes and its idiosyncrasies.

3.1 Introduction

The ready-mixed concrete (RMC) supply process is to be modelled in this research project. This process has been chosen, over other forms of producing concrete for use in construction (such as Volumetric Mobile Mixing), because it is the most commonly-used method (National Ready Mixed Concrete Association, 2001).

Before any introductions are made to the modelling methods considered in this project, it is important that the RMC supply process is introduced. In this chapter the procedures of the procurement, supply and use of RMC in major construction projects are discussed.

3.2 How RMC is procured in major construction projects

The procedure of procuring RMC for major construction projects is shown in figure 3.1. The design information (dimensions, quality specification, materials to be used, etc.) is passed from the designer to the construction contractor. One of the earlier tasks is to determine who should supply the project with RMC. The RMC needs to be fresh, and thus the distance from a project site to the concrete production plant must be minimal. Beyond the travel distance restriction, the factors affecting choice of supplier are cost, resource requirements and the ability to supply RMC at the required quality and delivery rates – otherwise in a traditional project there is no control (Pullen et al., 2000). The contractors must act quickly in communicating their requirements to materials suppliers, because of the lead-times involved in acquiring such materials; even though RMC is made on the day it is required, there is still a significant lead-time of roughly 4-5 days such in the modern-day demand placed on RMC producers by the construction industry.

The details which must be provided to RMC suppliers are:

- The concrete specifications (strength required, etc.).
- The proposed volume of RMC.
- The rate at which RMC should arrive (currently calculated from experience).
- The placing method to be used (e.g. truck-mounted pump).
- The number of truck mixers required for an operation.

After these details have been provided a RMC supplier will then provide a quote for price, delivery information and the concrete mix design details. These quotes should be compared for a number of local RMC suppliers and the most favourable supplier should be chosen.

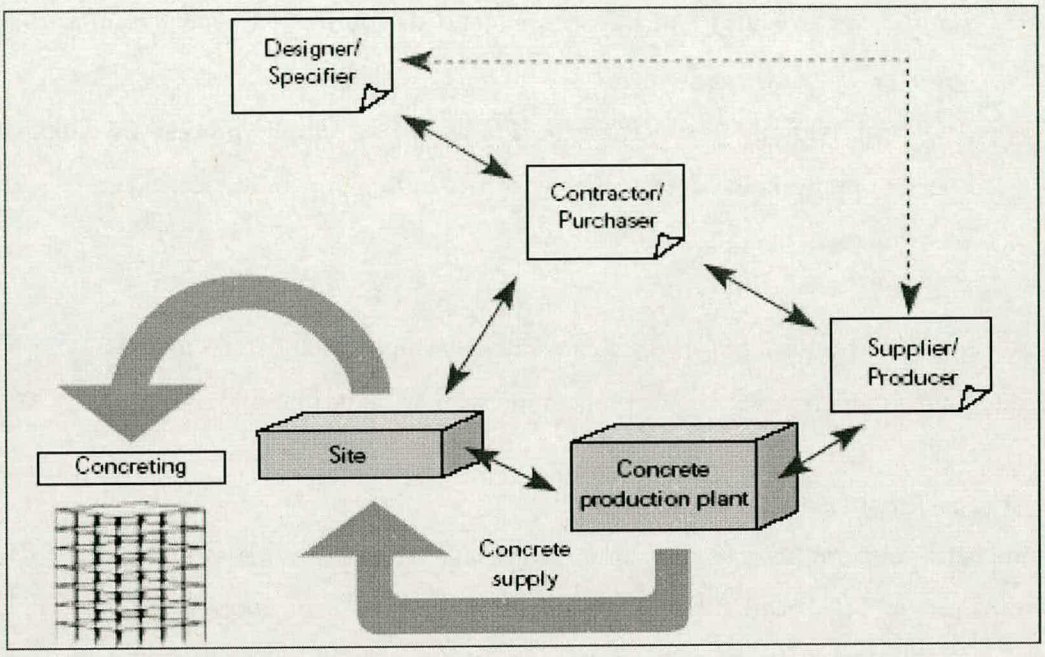


Figure 3.1: Procedure of RMC procurement (Pullen et al., 2000)

Finally, on the day preceding (or sometimes on the day of) the construction operation involving RMC, it is necessary for the contractor to confirm the order for RMC and provide additional information about site conditions, access points, etc. In theory, if the procurement procedure outlined here were followed 'to the letter' RMC would arrive at a project, in the correct form (quality and quantity), at designated intervals; this should allow the plant and operatives to be optimally managed.

3.3 The RMC supply and utilisation processes

This project is only concerned with modelling the RMC supply process when the method of transporting RMC around construction projects is a mobile, truck-mounted concrete pump. The reasons for this are three-fold:

- It is a classic example of client/server system, where the clients in this instance are the truck mixers and the server is the concrete pump. This type of system has been modelled successfully in the past using methods such as simulation, and this breadth of experience will hopefully aid the modelling process.

- A mobile concrete pump is the most widely used method of placing RMC in construction projects. Thus, any solutions presented in this thesis shall have a wider reach than if the model was based upon a less commonly used transportation method, such as a crane and skip, direct placement from a truck mixer or using a wheelbarrow.
- The final reason for concentrating on the RMC supply process involving a concrete pump is the data available for use in this project was collected through observation of operations of this type.

The RMC supply and utilisation processes are shown in figure 3.2. Before any RMC can be supplied to a construction project, some preparation work must be undertaken.

3.3.1 *Typical Preparation Work*

Before the batching plant has even begun to manufacture the RMC, there is some work to be done on the project site. Formwork must be placed in appropriate positions to provide stable boundaries for the RMC to be placed into. Additionally, the reinforcement steel – which gives a structure most of its strength – must be fixed into place. The RMC will be placed on top of this steel to form a reinforced concrete structure.

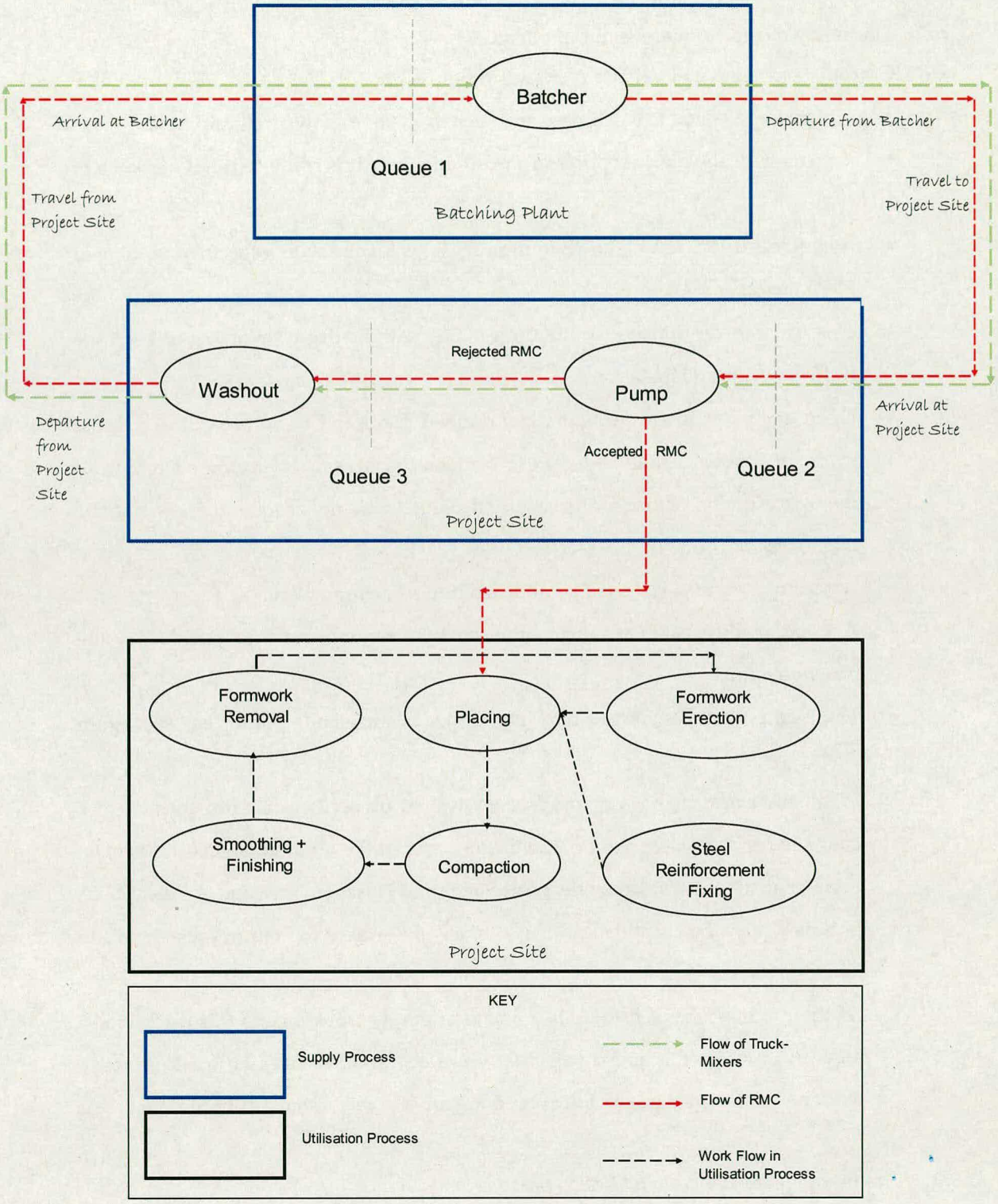


Figure 3.2: The RMC supply and utilisation processes

3.3.2 *The ready-mixed concrete supply process*

The RMC supply process is cyclical and is shown in figure 3.2:

- It begins at a concrete batching plant, such as the one shown in figure 3.3a.
- A concrete truck mixer is put into a position where it is ready to receive a batch of RMC.
- A concrete truck mixer (shown in figure 3.3a) is loaded with a specified volume of RMC mixture.
- This truck mixer departs for the project site, with its drum revolving to keep the RMC fresh and pliable.
- Upon arrival at the project site, if no queue is present a sample of the original truck mixers RMC shall be tested to confirm that it is of appropriate strength. If there is a queue of truck mixers on site, this truck mixer joins the back of the queue (see figure 3.3b).
- A poor quality material shall be returned to the batching plant.
- A good quality batch results in the truck mixer moving into position at the concrete pump.
- The load is then discharged into the hopper of a concrete pump (e.g. see figure 3.3c).
- When the truck mixer is relieved of its load, it proceeds to an area to have any concrete residue washed out of the drum – termed the washout area. If there is no need to queue the truck mixer is positioned in the washout area and is cleaned.
- Subsequently, the empty, clean truck mixer undertakes the return journey to the batching plant.
- At the batching plant, if there is a queue of empty truck mixers this truck mixer joins the back of the queue (e.g. see figure 3.3d). Otherwise, the truck mixer is loaded with another batch of RMC and the process re-commences once more.

3.3.3 *The ready-mixed concrete utilisation process*

The RMC utilisation process is as follows:

- The quantity of RMC that is discharged into the hopper of a pump is placed (distributed) in an appropriate area of the structure (see figure 3.3e).
- The RMC is then compacted (see figure 3.3f).

- After a suitable period of time, the concrete is then smoothed and finished.

Finally, once the RMC has set the wooden formwork is removed revealing a finished structure and this completes the utilisation process of RMC.

3.3.4 *Assumptions made in defining the above processes*

The following are some of the assumptions made in defining the RMC supply and utilisation processes:

- The supply chain of the materials required to form RMC has been ignored.
- The supply chains for the reinforcement steel (rebar), formwork and any scaffolding required have been ignored.
- Indirect costs have been ignored.
- Any need for the driver of a truck mixer to exit the vehicle has been ignored. This commonly occurs at both the batching plant and site, and is assumed to have been absorbed in the queues 1 and 2 in the above definition.
- Any cleaning (washout) of the truck mixer performed at the batching plant has been ignored. This is because this cleaning is not performed consistently and is therefore difficult to model.

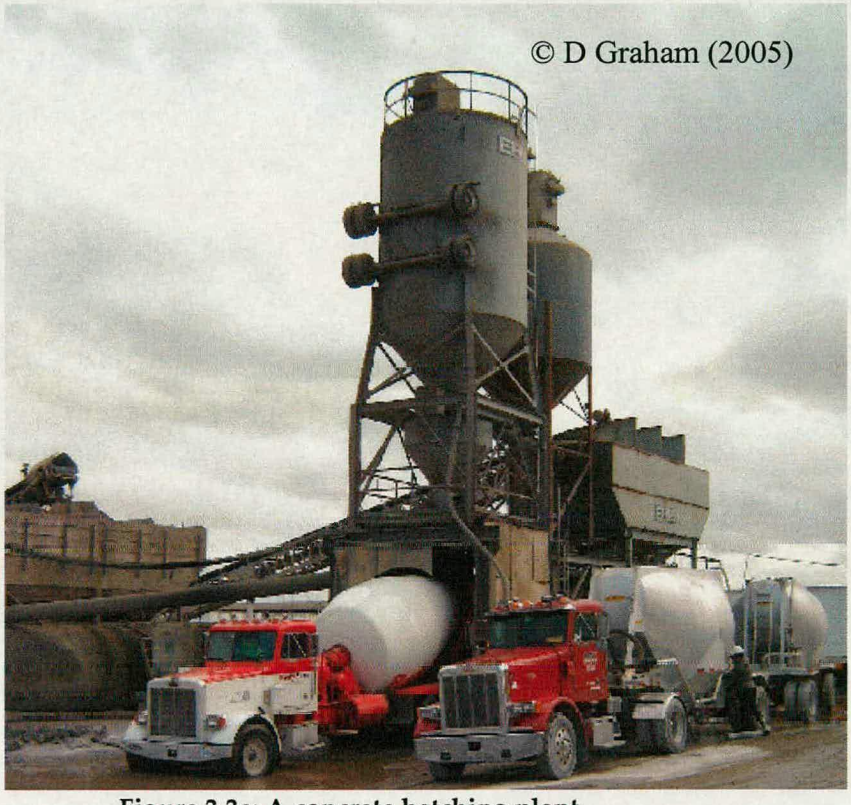


Figure 3.3a: A concrete batching plant

Figure 3.3b: A truck mixer queuing on a project site

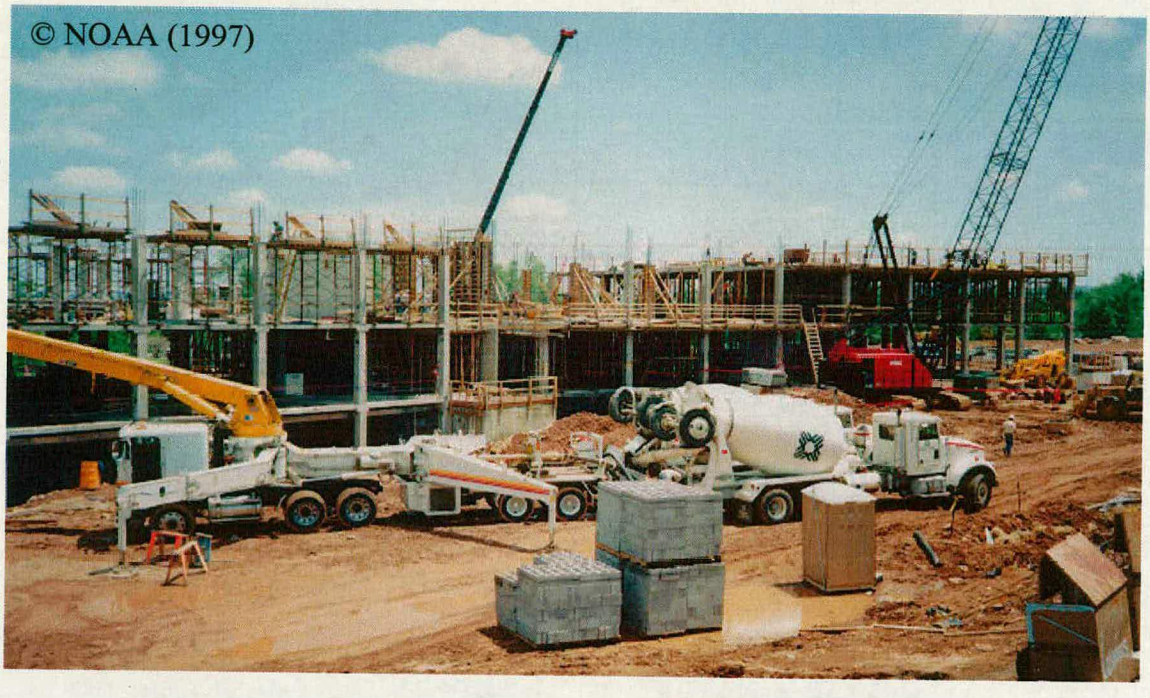


Figure 3.3c: A truck mixer discharging its load into a concrete pump

Figure 3.3d: Truck mixers queuing at a batching plant

Figure 3.3e: RMC being placed on a project site

Figure 3.3f: RMC being compacted on a project site

3.4 Chapter conclusions

In this chapter the procedures for the procurement, supply and use of RMC in major construction projects were discussed. The conclusions that may be drawn are:

- The details that a contractor (customer) must provide to an RMC supplier are the concrete specifications, the pour volume, the required rate of delivery of RMC, the placement method to be used and the number of truck mixers required.
- The lead-time for RMC is roughly 4 days.
- In this project, only concreting operations that had a mobile, truck-mounted pump will be modelled. This is because:
- This set-up provides a good example of a client/server system.
- It is the most widely used method of placing concrete.
- There is a data set available, recorded from observations that utilised a pump of this type.

CHAPTER 4

THE COLLECTION OF DATA

Data are the essence of model development projects, never more so than for those projects of a high practical nature. Without data models cannot be built nor tested and also they cannot be verified and validated as being true representations of reality. The careful consideration of the data which is collected, and the way in which it has been collected, is therefore the theme of this chapter.

4.1 Introduction

In order to substantiate the findings from the models produced in this thesis and to maximise the impact such findings have on construction practitioners and researchers, the data used in the development of these models was collected from real construction projects. The nature of the project on which this thesis is based was such that the vast majority of the data was not collected directly but by Dr Paul Dunlop, a previous PhD student of the University of Edinburgh, and acquired by Dr Simon Smith, an academic at the University of Edinburgh from a contractor's records. During the project the author acquired additional project data and spent time observing several cyclical, stochastic construction processes. All relevant details regarding data collection shall be discussed in this chapter.

Chapter four is structured as follows:

- **Section Two**. An examination of the available data collection methods.
- **Section Three**. An introduction to the method used to collect the initial project data is presented. Also the variables from this data set are listed.
- **Section Four**. The sources and dimensions of the original data set used for modelling are introduced.
- **Section Five**. Evolution of the data sets throughout the project. The details of the data acquired after this PhD project was commenced are discussed, including that used to bolster the original data set, and data for use only in the model validation process.
- **Section Six**. Chapter conclusions.

4.2 Available methods of data collection

There are numerous methods of collecting data for use in modelling. Flood and Kartam (1994) list some of the methods used in previous construction management and civil engineering problems. They are:

- **Observation of the process with a specific goal in mind**. For example a construction process could be observed with a goal of producing a model of the process productivity. Observation of construction projects is a time consuming enterprise and can constitute a large part of a research project (Kartam, 1996). However, this method is worth the effort, as it allows an accurate picture of a process to be developed in the data, and the researcher to gain a greater

understanding of the process (Law and Kelton, 2000). This method of data collection has been used in the past by Smith (1995) on the earthmoving process and Dunlop (2000) on the RMC supply and utilisation processes.

One drawback of this method is that once a process has been observed, it is difficult to expand the interest of the study beyond those variables collected in the first instance.

Although this would most likely have been the best method of collecting the data to be used in this project, it was not possible due to a lack of resources. Rather, the next method of data collection has provided all of the project data.

- **The use of historic data**. Historic data is viewed in this thesis as having been collected in two ways: a specific study of a process, undertaken for the purposes of this thesis - that is to produce models of the process itself; or a more general study of a process, with a broad aim of learning more about the process. The former is likely to produce data which is more pertinent to this project than the latter.

  There is a general lack of historical data on construction processes, mainly due to commercial sensitivities. This notwithstanding, there have been two studies of note which used historical data: Karshenas and Xin (1992) used historic data specifically collected for the purposes of modelling the productivity of earthmoving equipment; Smith (1998) used data from a generic study (second method from above) of the construction of a motorway viaduct to model concreting processes.

  As stated above, historic data has provided the data for use in this project. The data sets produced are a combination of data collected with the specific purpose of modelling a process (see 4.3 – 4.4) and generically collected data (see 4.5).

- **By analysing the output of another model of the same process** and using this data, otherwise known as meta-modelling. Gagarine et al. (1994) used this method to undertake a finite element analysis of a bridge, giving different loading

conditions. The resulting strains from the finite element analysis were used as the inputs to a neural network model of the bridge.

This method of data collection was not considered in this project as there is some likelihood that the model producing the output of a process contains significant errors. This would result in further, probably more significant errors being produced by any subsequent models of a process.

- **Through conducting a questionnaire survey** of construction practitioners, academics and other interested parties. This is the most common method of data collection in construction management research (Fellows and Liu, 2003).
- **Via consultation with an expert of the process** to be modelled. Both this and previous collection method are more pertinent to 'soft' subjects. In this project these two methods were not considered because a computer-based solution is sought, and it would be difficult to produce such a solution using the mainly qualitative-type data produced by this (and the previous) method.

4.3 Method of collection of the original data sets

The original data sets were collected by Dunlop (2000) in his observational study of the RMC supply process. In this project, these data sets are viewed as historical data (as they were not collected by the author) which were collected with a specific purpose in mind – to model the process. As this is a shared aim of Dunlop's project and this one, these data sets should be very applicable in this project. The methods of observation of the process shall now be discussed.

The focus of Dunlop's work was to investigate the application of modelling methods to solve the problems of the RMC supply process (Dunlop, 2005). Dunlop's work differs from that undertaken in this research project for three main, and important, reasons:

- Dunlop (2005) decided to focus upon developing a linear-type model of the RMC supply process. This type of model was considered in this research project, and while found to be capable of modelling the RMC supply process, it was thought that a non-linear, stochastic solution would be more capable of accurately modelling the process (see chapter 5), and thus such a solution was considered.
- This research project placed a focus on developing a model that was capable of being practically implemented (see chapter 5). While this was also considered by

Dunlop (2005), the issue was not considered as important in his work; proving the technological capability to model accurately was.

- Finally, Dunlop (2005) centred all of his research on the issues of concreting, while in this research project there has been an attempt to look beyond concreting to the application of the methods developed (in chapters 6 and 7) to other stochastic, cyclical processes in construction (see chapter 10). Dunlop's work provides a very detailed explanation of the concreting process and the issues surrounding the concreting process, while this research project provides a broader overview of many processes.

The process of RMC supply was introduced in chapter 3, where it was pronounced to be cyclical. As can be seen in figure 4.1, this is because a truck mixer begins and ends its role in the supply process at a concrete batching plant, forming a cycle. In all of the projects studied by Dunlop (Dunlop and Smith, 2003) more than one truck mixer was involved, meaning that Dunlop could not simply travel with any particular truck mixer to record its performance. This method would have resulted in not observing the performance of all the other truck mixers, and a very skewed set of results based only on one truck mixer. A difficult situation ensued in which information needed to be collected from two places at the same time (batching plant and construction site). This difficulty was resolved by observing the performance of truck mixers, at the concrete pump and washout area on the construction site. This involved making a time measurement using a stop-watch and recording these following measurements on a standard plant observation form: the project being observed; the date of the pour; the location on the structure of the pour; the type of pour e.g. a base pour; the truck mixer reference number; the time of arrival of a truck-mixer on site; the time a truck mixer began to discharge its load; the time this discharge was completed; the time at which the truck-mixer began to be cleaned out, and finally when this cleaning process finished. Concurrent to Dunlop's observations, the truck mixer operators compiled a record of: the travel distance; the time of their arrival at the batching plant; the start time of batching (loading) their truck mixer; the time batching was complete, and the volume batched in their truck-mixer. In addition, the construction manager of each project provided the concrete mix type.

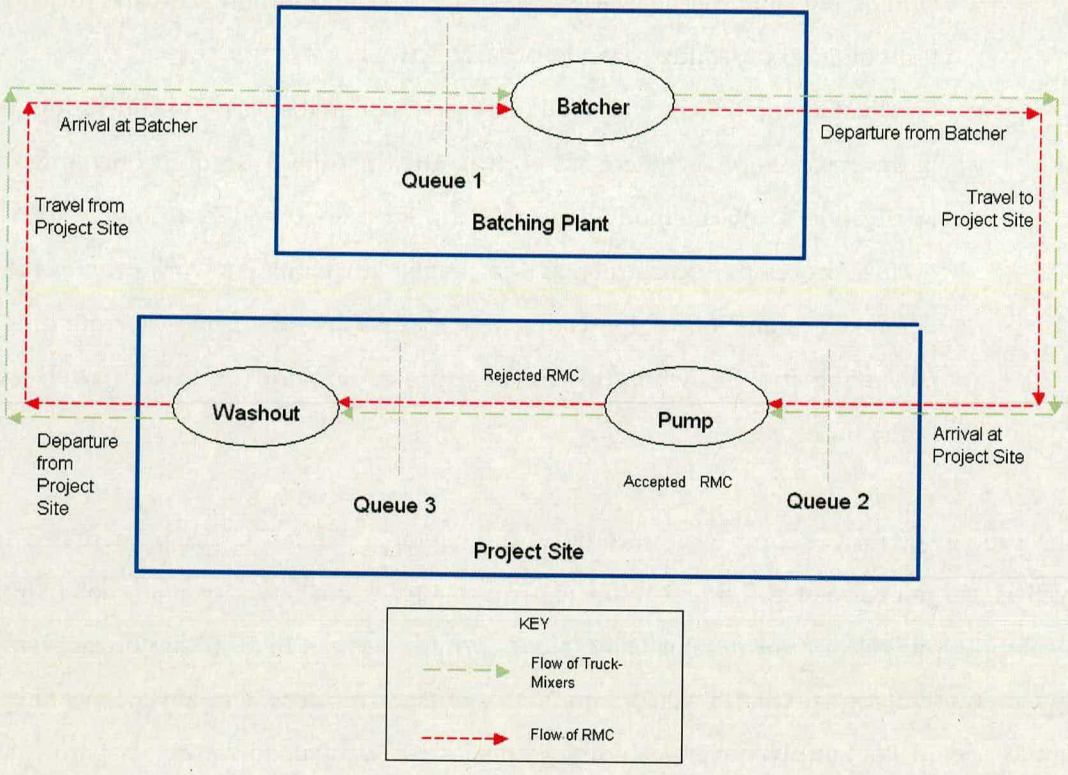


Figure 4.1: The RMC supply process

The above information for each pour was collated using a standardised spreadsheet, such as that shown in figure 4.2. A number of values of interest were calculated for each concrete pour from these collated spreadsheets. These values were:

- The average volume of concrete per truck mixer.
- The number of truck mixers involved in the pour.
- The total volume of the pour; pour duration.
- The time taken to batch each truck mixer, known as truck batch time.
- The average time taken for each truck mixer to travel from the batcher to site or vice-versa (average travel time).
- The time taken to position a truck mixer at the concrete pump (position time).
- The time taken to discharge the load of a truck mixer (discharge time).
- The time taken to wash a truck mixer (washout time).
- The percentage amount of time the pump was in use (pump utilisation rate).
- The percentage amount of time the collective group of truck-mixers were in use (truck mixer utilisation rate).

- Lastly, the productivity of the pour (total volume of pour/pour duration).

This process of observation produced a large set of variables (28 in total), whether raw or derived from other pieces of information. Later (in chapter 5), the amount of variables to be used in model development shall be reduced to those which hold statistical significance and can be practically determined by a construction practitioner at the time of placing an order for RMC.

While a large number of variables were collected from observation of the RMC supply process, there are numerous variables which were excluded from the collection process, and hence excluded from the model itself. These excluded variables can be broadly placed in the following categories: personnel issues; equipment issues; and, materials issues. The following is a description of some of the excluded variables by category:

Personnel Issues: size and quality of 'pool' of available workers; age of workers; experience of workers; training needs of workers – including time spent 'on the job' training; over time requirements; experiential learning of workers – work improvements due to familiarity with the construction site, etc.; sickness rates; accidents and injuries; health and safety of workers, the general public, etc.; warm-up period for workers beginning work; relationship between contractor and supplier – can affect the quantity of truck mixers provided to a contractor, etc.; availability of batch plant workers to produce RMC.

Equipment issues: age, quality and capital cost of equipment; maintenance history; breakdown rates of batch plant, truck mixers, etc.

Materials issues: availability of raw materials for use in RMC production.

These variables, while not being explicitly included in modelling, are not completely ignored. This is because the impact of each of these variables is reflected in the statistics used to build the simulation model. For instance, if the workers are inexperienced then the statistical distribution for travel time, position time and discharge time would reflect that inexperience.

Project Description	Wastewater treatment tank at Fraserburgh
Pour Date	19th July 2000
Location of Pour on Structure	Outside Walls of Tank Structure
Type of Pour	Wall
Travel Distance from Batching Plant to Site	12.6 miles
Concrete Mix Type	C40

Truck Mixer ID	Batch Volume (m3)	Time of Arrival at Batching Plant	Time Batching Started	Time Batching Finished	Time of Arrival at Site	Time Discharge Started	Time Discharge Finished	Time Washout Started	Time Washout Finished
N584	8	11:00	11:05	11:10	11:25	11:31	11:40	11:41	11:45
T145	8	11:00	11:11	11:20	11:40	11:44	11:53	11:55	11:58
W815	8	11:30	11:35	11:40	11:55	12:00	12:07	12:08	12:14
N584	8	12:07	12:08	12:15	12:32	12:34	12:40	12:42	12:48
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

Figure 4.2: Example of information collated from one observed concrete pour (Dunlop, 2000)

4.4 The sources and dimensions of the original data sets

The original data was collected using the time-study method outlined above from four major construction projects in the East of Scotland, United Kingdom. The following provides a description of each project:

- *Projects 1, 2 and 3: The Aberdeenshire wastewater treatment works*

The aim of this project was to design, build and operate (for 30 years) three wastewater treatment plants in Aberdeenshire, to serve more than 250,000 customers. The financing of this £80 million project was organised under the UK Government's Private Finance Initiative (PFI), and a concession company to provide this finance called Aberdeen Environmental Services Ltd was formed. This company consisted of Balfour Beatty Plc., Kelda Group and Tyco Tech Ltd – all experienced with this type of finance and development. The main contractors for the project were Balfour Beatty Construction (construction contractor) and Earth Tech (geotechnical contractor).

The three plants are located in Nigg, Peterhead and Fraserburgh (see figure 4.3) and for the purposes of this thesis they are all considered as individual projects; each one has unique characteristics.

From the Aberdeen project, 154 concrete pours were observed. The volume of these pours ranged from 12 to 294m^3, with an average of 57 m^3. Three types of concrete pour were observed in the following proportions (to the total 154 pours): base 36%, wall 63% and column 1%.

From the Fraserburgh project 20 concrete pours were observed. The volume of these pours varied from 32 to 315 m^3, with an average of 126 m^3. The amount of base and wall pours observed were almost equal (base 50% and wall 44%), while the number of column pours observed constituted only 1% of the total.

From the Peterhead project, 10 concrete pours were observed. The volume of these samples ranged from 18 to 294 m^3, with an average of 143 m^3. There were no column pours observed in this project. The majority of pours (80%) observed were of the base-type, the remainder being wall pours.

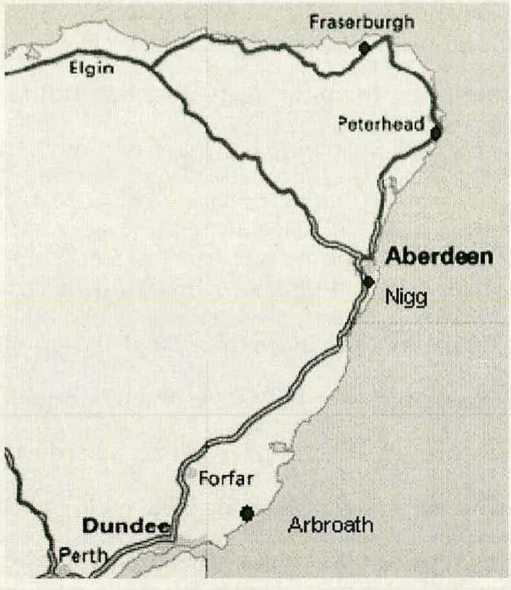


Figure 4.3: The locations of East of Scotland construction projects (after, Collins Bartholomew Ltd., 2003)

- *Project 4: The Tayside wastewater treatment works*

Like the above projects, this one involved the construction of a wastewater treatment plant, in Arbroath, near Dundee (see figure 4.3) using the UK Government's Private Finance Initiative (PFI) to provide the finance required. The aim of this project, however, was significantly different to those outlined above as the group who were to selected by the UK Government to commission the project were also required to design, procure, construct and performance test the treatment works. The consortium selected for the role comprised of Bechtel Water Ltd and Morrison Construction Ltd. As stated above, due to the nature of the contract of this project, these two companies were working together acting as the construction contractor.

48 concrete pours were observed on the Arbroath project, measuring from 12 to 180 m^3, with an average of 76 m^3. The amount of base pours witnessed in this project (58% of the total) outweighed the number of wall pours (42%). No column pours were observed.

To summarise, a total of 232 concrete pours were observed from real construction projects, resulting in a large set of data being available for use in this thesis. When compared with the data sets of previous published work in this (prediction and estimation) area of construction management research, the data set used in this project can be considered significantly larger; Rowings and Sonmez (1996; 1998) used 112 observations of concrete pours to model labour

productivity using both linear regression analysis and neural networks; Abourizk et al. (2001) used 39 observations of pipe-installation activities to develop a 2-stage neural network model of labour productivity rates; Arditi and Tokdemir (1999) used 102 cases taken from the Illinois circuit court to compare the capabilities of neural networks and case-based reasoning to predict the outcome of construction litigation.

Additionally, the data is spread over a large range in terms of pour type and pour volume (12 m^3 to 315 m^3).

4.5 Evolution of the project data

The data set detailed in 4.4 did not remain static as this thesis progressed. Rather, after undertaking the literature review it was determined that the most significant variable that was ignored in this initial collection of data was impact of the weather. This was determined from the work of Rowings and Sonmez (1998), who quantified the significance of the impact of weather in construction projects. It was deemed likely that the weather could adversely affect the rate at which concrete is used, and hence the rate at which it can be delivered. Thus, the ignorance of weather related variables could reduce the ability of any models to reproduce the RMC supply process.

To address the deficiency in the data set, additional data was procured from the UK Met Office, which holds historical weather records from most parts of the United Kingdom. This data is viewed as historical data which has been collected for general use (see 4.2). The variables contained in this additional data were as follows: the daily maximum temperature (°C), the daily minimum temperature (°C), the amount of rain (mm), the amount of sunshine (hours), the grass minimum temperature (°C), and the wind speed (knots). These variables at first glance appear to be impossible to predict, but the Met Office provides such information up to 5 days in advance, removing the need for any guesses to be made. Thus, the information pertaining to these variables was introduced to the original data set, resulting in a total of 34 variables available for use in modelling.

Further concerns emerged regarding the original data set. One subject of trepidation was the fact that all of the data was observed on projects located within close proximity to one another. It was felt that this could result in models of the RMC supply process being produced that

would be pertinent only to the East of Scotland. Another matter of concern was that only wastewater treatment works were being constructed in the observed projects. This potentially could have resulted in models that can only reproduce RMC supply for such structures – this thesis aims to be much broader than that. Thus, an additional data set was obtained that contained observations of concrete pours in a viaduct over a major motorway in the north west of England (see figure 4.4). Smith (1998) acquired this data set, which was originally collected by a contractor in general observations of concrete pours; thus this data is viewed as historical data which has been collected for general use (see 4.2).

The data contained information from 50 concrete pours, ranging in volume from 51 to 470 m^3, with an average of 210 m^3. This new data set also reduced the final concern, which was that the original data could restrict a model's ability to reproduce concrete pours, which were not of the base or wall type. The data collected by Smith contained a number of column pours (15% of the total) and thus should increase any model's understanding of the rate of supply of RMC in this type of pour.

Figure 4.4: The location of the motorway viaduct project

(after, Collins Bartholomew Ltd., 2003)

To gain a measure of the ability of any models to reproduce the RMC supply process, it was determined that Smith's data set acquired from should be retained for use in model validation. The premise being, if a model based only on the original data can reproduce concrete pours

which: have not been used in the model development; are from a different part of the UK; are from a project of a different type; are of a broader range of pour volume; and have more varying pour types, then that model can be declared as valid as possible without being employed in a real-life construction project. This has not been possible due to the time constraints placed upon this thesis. However, it is a suggestion for future work in this area of research.

4.6 Chapter conclusions

This chapter contains the details of the sources of the project data were provided. In summary:

- The original data sets comprised historical data collected from real concreting operations by Dunlop (2000), producing a data set of 232 pours with 28 variables. The observed projects involved wastewater treatment works, being constructed in the east of Scotland. The pours were mainly of a base and wall type.
- Weather was thought to be an important element and thus additional information on the weather witnessed during operations was collected from the Met Office, UK. This data increased the number of variables from 28 to 34.
- To help make the data set more broad an additional data set was acquired which was from a very different type of project (viaduct construction), undertaken in a different location (England) and comprised other pour types (e.g. column). This data set added 50 operations to the collection. It was determined that the additional data should be retained for use in model validation, not in any model development capacity.

CHAPTER 5

STATISTICAL ANALYSIS OF THE PROCESS DATA

Data cannot be simply taken and used, and conclusions drawn. In this research project, especially with data collected from elsewhere, a decision was taken to ensure the data was fit for the purpose.

5.1 Introduction

A statistical analysis of a data set, when undertaken in a robust manner can at best provide telling evidence about which factors are key in driving the particular system or process being studied, and at worst provide delicate nuances about the same key factors. This thesis has been approached with an air of optimism and thus it was hoped that the key factors or variables of the RMC supply process could be defined through a robust statistical analysis. The following outlines a systematic, robust statistical analysis performed on the RMC supply process data:

- **Section Two**. Before the key variables could be identified, the variables that were collected from the construction projects outlined in the previous chapter are described.
- **Section Three**. Subsequently, an examination was undertaken to determine if the collected variables were normally distributed – because if they were not, the use of standard statistical techniques such as linear regression would have been hampered in this analysis. As it happens, the collected data was, on the whole, normally distributed, thus lending itself to standard statistical methods of analysis.
- **Section Four**. The data set contained non-numerical and missing data and potentially held outliers. Before an adequate model could be made, these issues needed to be resolved, and this was done using a variety of methods, one of which was a new technique developed in this research project for identifying outliers in construction data sets. Resolving these issues resulted in a data set, which was of a standard numeric form, with no missing data, and no outliers that could skew any models estimates of the RMC supply process. This was the final form of the collected data, and it was then possible to identify the key variables of the process from this 'complete' data set.
- **Section Five**. The next precursor to commencing any model development was to ensure that a non-linear modelling method, such as simulation, was apposite for this process. To ascertain this, a linear regression analysis was performed on the collected variables – the thought being if a linear model could predict the performance of the RMC supply process (e.g. productivity) adequately, there is no need to complicate matters by modelling the process non-linearly; thus adhering to the principle of parsimony. The conclusion was, however, a non-linear model is required to estimate the RMC supply process - thus the key variables of this process must still be defined for use in the development of such a model.

- **Section Six**. The overall purpose of this thesis is to produce a practical model capable of estimating the performance of the RMC supply process. A model will only be practical if it is built upon variables that can be known by construction practitioners. To ensure that a practical model was developed in this thesis a framework for model development was devised and is detailed in this section. Subsequently, the variables upon which the models developed in this thesis are based are identified.
- **Section Seven**. Finally, a summary of the findings of this statistical analysis is presented and conclusions are drawn.

5.2 The variables contained in the collected data

Preceding a statistical study of the RMC supply process data it is important to define the variables available in the collected data set and to clarify the nature of these variables. A list of the collected variables is shown in table 5.1 and they have been separated into two distinct groups: response or dependent; and, explanatory or independent. A simplified description of the relationship between these two variable categories is to say the response of the process is governed by a combination of the weighted values of the explanatory variables. This is the form any model of the RMC supply process will take – the user provides a value for each explanatory variable and the model shall provide a response to these values.

In the collected data, there are 26 explanatory variables which are suspected of holding a relationship with the 8 dependent variables of the process. Note that month and date of pour are one and the same and are interchangeable in this discussion.

Variable	Units	Variable Type	Data Type	No. of Missing Data Points
Project	-	Independent	Nominal	0
Date (Month) of Pour	-	Independent	Ordinal	0
Travel Distance	miles	Independent	Numerical	0
Location of Pour on Structure	-	Independent	Nominal	0
Type of Pour	-	Independent	Nominal	0
Concrete Mix Type	-	Independent	Nominal	45
Number of Truck-Mixers	-	Independent	Numerical	0
Total Pour Volume	m^3	Independent	Numerical	0
Truck-Mixer Reference Number	-	Independent	Nominal	0
Batch Volume	m^3	Independent	Numerical	0
Average Volume / Load	m^3	Independent	Numerical	0
Time of Arrival at Batching Plant	hours, minutes	Independent	Numerical	0
Start Batching Time	hours, minutes	Independent	Numerical	0
Complete Batching Time	hours, minutes	Independent	Numerical	0
Time of Arrival on Site	hours, minutes	Independent	Numerical	0
Time Discharge Began	hours, minutes	Independent	Numerical	0
Time Discharge was Complete	hours, minutes	Independent	Numerical	0
Start Time of Washout	hours, minutes	Independent	Numerical	0
Completion Time of Washout	hours, minutes	Independent	Numerical	0
Pour Duration	minutes	Independent	Numerical	0
Daily Maximum Temperature	°C	Independent	Numerical	0
Daily Minimum Temperature	°C	Independent	Numerical	0
Amount of Rain	mm	Independent	Numerical	0
Amount of Sunshine	hours	Independent	Numerical	0
Wind Speed	knots	Independent	Numerical	23
Grass Minimum Temperature	°C	Independent	Numerical	0
Batch Time	minutes	Dependent	Numerical	0
Travel Time	minutes	Dependent	Numerical	0
Position Time of Truck-Mixer	minutes	Dependent	Numerical	0
Discharge Time of Truck-Mixer	minutes	Dependent	Numerical	0
Washout Time	minutes	Dependent	Numerical	0
Pump Utilisation	%	Dependent	Numerical	0
Truck-Mixer Utilisation	%	Dependent	Numerical	0
Process Productivity	m^3/hour	Dependent	Numerical	0

Table 5.1: The variables contained in the RMC supply process data set

5.3 The underlying distributions contained in the data

The approach of most people, for better or worse, seems to be to regard data as being normally distributed unless there is some reason to believe that this is not true (Manly, 1994). However, knowing the underlying distribution of a set of data is important because a number of commonly used statistical analysis methods, such as linear (LS) regression make the assumption that the data is normal (Draper and Smith, 1998). To gain valuable information about the RMC supply process this study intends to use LS regression and therefore, it is necessary to make an assessment of the distribution of the data.

5.3.1 A method for calculating the distribution of the data

The collected data set is multivariate – there is more than one variable involved and the most basic method of calculating the multivariate distribution of data is to determine the distributions that represent the individual variables, and form a cumulative judgement from these (Manly, 1994). This simple practice has been implemented here, the details of which are as follows:

- **Apply theoretical distributions to each of the variables in table 5.1.** A bespoke statistics package for construction data was used in this thesis. The package, *PDFit*, was developed by the author in a previous Honours thesis and is discussed in detail in Graham et al. (2005). *PDFit* was used to apply 27 probability distributions to the data (see 6.5 for details) available for each of the 34 collected variables, and to assess which of these distributions provided the best representation of each.
- **Assessment of the representation that the probability distributions provided for each of the variables.** To determine the best representative for each variable two goodness-of-fit tests, the Anderson-Darling (AD) test and the Kolmogorov-Smirnov (KS) test (see Law and Kelton, 2000), were performed on the data by PDFit.
- **Assessment of the underlying distribution of the complete data set.** The individual results for each variable were combined to build up a picture of the most representative distribution of the whole data set, and thus the process itself.

5.3.2 Distribution analysis results

The results provided by the AD test for the RMC supply process have been ranked, and the percentage of times that a particular distribution was ranked as providing the best representation for a variable is shown in figure 5.1.

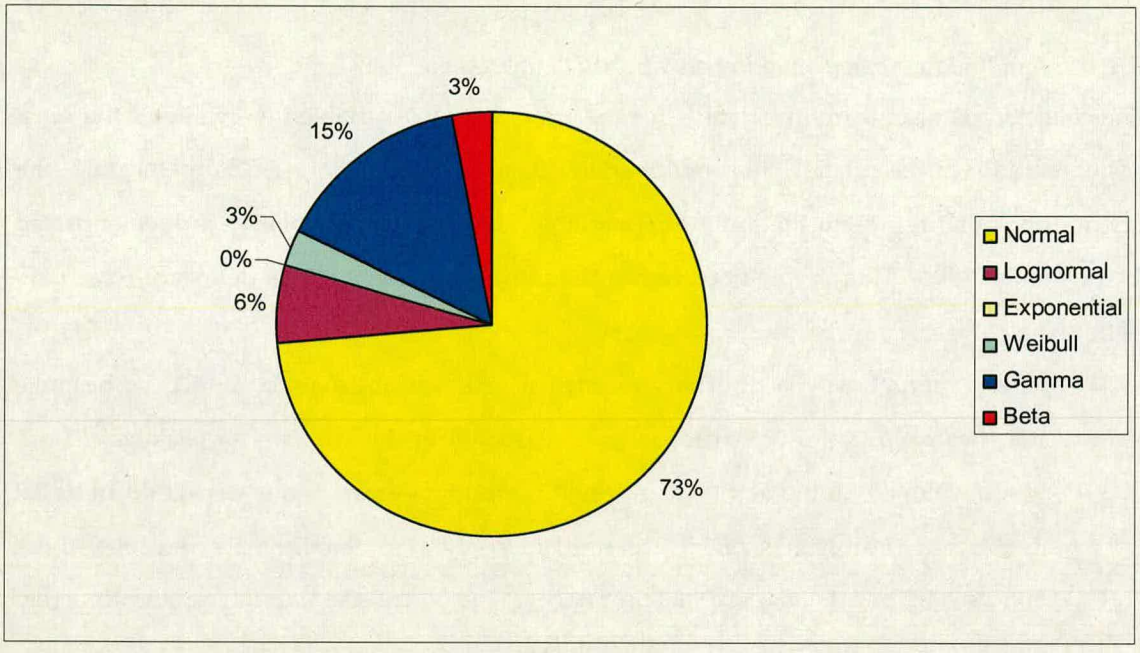


Figure 5.1: %age of times the distributions were ranked as most representative

For the variables: project; number of truck-mixers; truck-mixer reference; time of arrival at batcher; complete batching time; pour duration; daily minimum temperature; wind speed; and process productivity, non-normal distributions were found to provide the best representation. However, as can be seen in figure 5.1, the normal distribution does provide the best representation of 73% of the variables and it is possible to conclude that the overall data set is on the whole, normally distributed. Support for this conclusion was also provided by the results of the KS test, which found that: the normal distribution was representative of 74% of the whole data set; the gamma distribution represented 14% and the lognormal represented 7% of the whole data set.

Now that the data set has been confirmed as normally distributed standard statistical techniques may be used.

5.4 Dealing with problems contained in the data and reducing the variables

The main problems contained within the data set to be dealt with were how to deal with missing data, outliers and non-numerical data values. An additional issue concerns the ratio of the amount of data collected to the number of process variables is small (232:34), possibly resulting in a data set that is sparsely populated. This sparseness can result in problems when

modelling the RMC supply process, such as an inability for a model to be able to predict the RMC supply process performance measures in future projects. To help avoid data sparseness, the correlated variables were identified and removed.

5.4.1 Dealing with missing values

One of the most important issues to be addressed in this statistical analysis is how to deal with missing values in the data set. In the collected data set, the wind speed and concrete mix variables have missing values (see table 5.1). Standard statistical methods e.g. to identify outliers, have been developed to only analyse rectangular data sets (Little and Rubin, 1987). Therefore, to utilise statistical methods to learn more about the RMC supply process, the data sets must be converted into a rectangular form. To achieve this end the following strategies were examined:

- **The removal of the variables which contain missing data values.** This will maximise the amount of the original data available for modelling, but may result in the loss of variables which are important in describing the system.
- **The removal of the cases which contain missing data values.** This is the opposite of the removal of variables. Variables essential to describing the process may be included, but with an increase in the sparseness of the data set.
- **The imputation of missing data values**, estimated using: Mean, LS Regression, Expectation-Maximisation (E-M), Linear Interpolation (Interpolation) and Linear Trend at Point (Trend) (see Little and Rubin (1987) for an explanation of these estimation methods). If the missing values can be accurately estimated using one of the above methods, this method would maximise the density of the data set, while including the maximum number of variables.

For each of the above methods a new data set was created and a linear (LS) regression analysis (using the step-wise method) was undertaken on each; the results of each are shown in table 5.2. The findings are that the adjusted R^2 value is highest when the values are estimated using linear interpolation. Taking into consideration that the imputation of missing values provides the maximum available data and provides the most accurate analysis of the system, it has been decided to use the linear interpolation method to remove the missing values and improve the data set.

Concreting Process
Only Positive Non-Numerical Variables

Significance Level %	Missing Values Strategy						
			Estimation Method				
	Variables Removed	Cases Removed	Mean	Regression	E-M	Interpolation	Trend
5	77(7)	81(7)	81(8)	81(8)	81(9)	82(8)	81(9)
1	76(8)	80(9)	81(10)	81(10)	80(11)	81(10)	81(10)

Notes:
The values are the adjusted R^2 values from LS regression analysis
The values in brackets are the number of steps required to reach the final adjusted R^2 value.

Table 5.2: Missing values LS regression results

5.4.2 *Dealing with outliers in the data*

This section aims to: explain what outliers are; discuss why outliers are important in the modelling exercise performed in this thesis; describe how they arise in data; discuss two methods of dealing with outliers; focus upon a new method for identifying outliers in general data sets and in the RMC supply process data set.

- *What are outliers?*

An outlier has been defined by Barnett and Lewis (1994) as an observation (or subset of observations) that appears to be inconsistent with the remainder of that set of data. There are two types of outlier: a data point that is genuinely an extremity of the data set; or, a data point that really belongs to a different distribution than the one thought to represent the data (this is known as a contaminant). Regardless of the type of outlier, it is characterised by its impact on the observer (not only will it appear extreme, but it will seem, in some sense, surprisingly extreme). Therefore, the detection of outliers in a data set is subjective, and this fact is highlighted in the above definition by the phrase 'appears to be inconsistent'. To illustrate the subjectivity of identifying outliers in a data set, an example of a suspected outlier in English marital law, taken from Barnett and Lewis (1994) is presented:

In 1949, in the case of Hadlum v. Hadlum, Mr Hadlum appealed against the failure of an earlier petition for divorce. His claim was based on alleged adultery by Mrs Hadlum, the evidence for which consisted of the fact that Mrs Hadlum gave birth to a child on 12 August 1945 which was 349 days later than Mr Hadlum had left the country to serve King and nation!

The average gestation period for the human female is 280 days: 349 days seemed to Mr Hadlum surprisingly long. He adjudged it an outlier. The court however did not, and Mr Hadlum was unsuccessful in his appeal. Statistically, the observation of 349 days is extreme: surprisingly so, and we designate it an outlier. But this does not resolve the matter of the status of the observation, 349. It could be a genuine observation from the distribution of gestation periods, which just happened to be surprisingly extreme. This is essentially what the court ruled. On the other hand it might be a contaminant: an observation from some other distribution (in this case with later origin) - this is precisely what Mr Hadlum was claiming! Thus illustrating the subjectivity of deciding if an observation is an outlier from the remainder of the data.

- *Why are outliers important in modelling the RMC supply process?*

Now that a definition of an outlier has been provided, the importance of outliers in modelling the RMC supply process shall be discussed. Outliers occur more commonly than is generally appreciated, a few percent of wrong values tending to be the norm (Seber, 1984; Hampel et. al, 1986). Therefore, the possibility of the presence of outliers in the data set cannot be ignored. Outliers in the collected data set would have an impact on the modelling of the process if using:

  - **LS regression analysis** (which is used in this project). Outliers in LS regression analysis are the data that have unusually large response errors (not large response values) and hence lie far from the majority of the data and the regression function (Myles, 1997; Draper and Smith, 1998). Their effect on LS regression is to pull the fit away from the trends suggested by the bulk of the data as figure 5.2 illustrates.

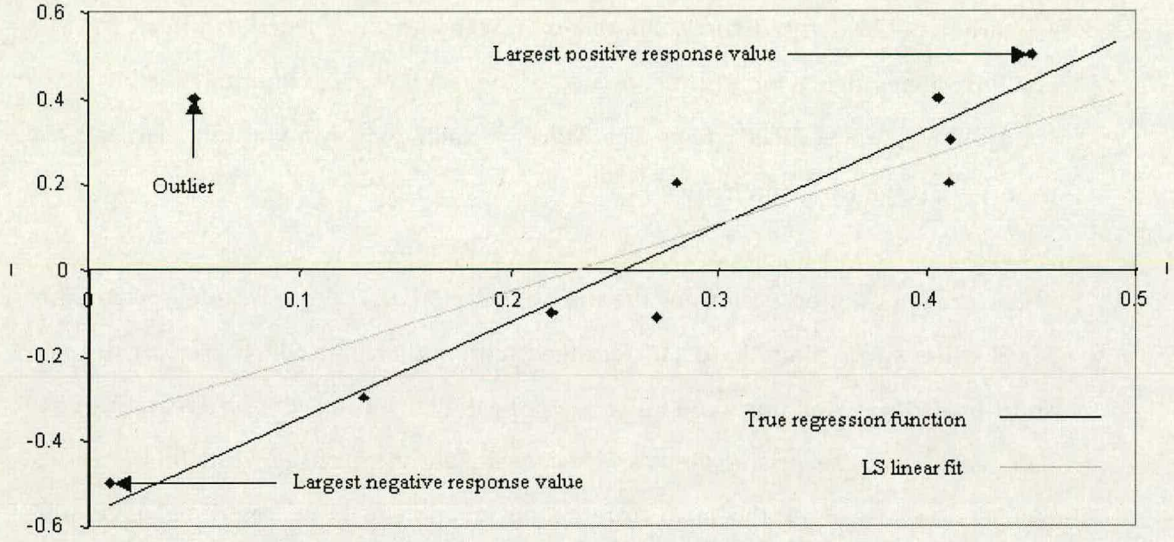


Figure 5.2: The effect of an outlier on linear (LS) regression

- **Simulation** (which is contemplated for use in this study). A simulation model is driven by estimates of the durations of activities contained within the process. These estimates are generally derived from probability distributions that represent the activities. The effect of outliers on probability distributions is to either extenuate the tail of the correct distribution, or point to the need for a more skewed distribution to correctly represent an activity duration. In the former case, there will be a greater probability of the distribution generating an extreme time (high or low), which clearly will have an impact on the output from a simulation model. In the latter, it is likely that a completely incorrect estimate of activity duration shall be produced – rendering any simulation model affected by outliers in this way useless.

- *How do outliers arise?*

In recording observations, different sources of variability that give rise to outliers can be encountered (Anscombe, 1960; Hawkins, 1980; Barnett and Lewis, 1994):

- **Inherent variability.** This is the expression of the way in which observations vary over a data set; such variation is a natural feature of a data set. This variability is uncontrollable and reflects the correct distributional properties underlying the data set.
- **Measurement error.** Often physical measurements are made of the variables in a process or system under study. Inadequacies in the measuring

instruments superimpose a further degree of variability on a variable. The rounding of measured values, or mistakes in recording act to compound the measurement error. Some control of this type of variability is possible.

- **Execution error.** A further source of variability arises in the imperfect collection of data. Inadvertently, a biased sample or individuals (examples) that are not truly representative of the data set may be included. Sensible precautions may reduce such variability.

Irrespective of the source of outliers, they afflict both LS regression and simulation modelling, and consequently outliers are of interest here. It was therefore essential to investigate methods for handling outliers.

- *Methods for handling outliers*

There are two widely used methods of handling outliers (Barnett and Lewis, 1994):

- **Identification and Removal of Outliers.** The detection of outliers in multivariate data - such as the process data set - is very difficult (Ripley, 1996; Barnett and Lewis, 1994; Hampel et. al, 1986). However, methods of identifying outliers in multivariate data do exist and it is yet uncertain as to the presence of outliers in the process data. Therefore, some methods for detecting outliers shall be employed to remove this uncertainty.
- **Apply Accommodation (or Robust) Methods to the Data.** Accommodation methods are resistant to the effects of outliers (Seber, 1984) and can be used to: estimate the parameters of probability distributions; fit a linear regression function to a data set; or, fit a MLP regression function to a set of data. A consequence of this resistance is that they do not require the identification of outliers (Hampel et. al, 1986). Thus, robust methods could have a direct influence on the modelling methods outlined above, and could save the data analyst considerable time and effort. Although clearly useful in certain circumstances, if outliers do not exist in the data, robust methods are superfluous. Therefore, these methods shall only be employed if the presence of outliers in the data set is confirmed in the following study.

- *Identification of outliers in the process data set*

The identification of outliers in multivariate data is usually based on the Mahalanobis distance of an observation from all other data points (Santos-Pereira and Pires, 2002). The

Mahalanobis distance is a multi-dimensional extension of standardised scores and is defined in equation 5.1 (notation can be found in appendix H).

$$D^2 = (x-\mu)' V^{-1} (x-\mu) \quad (5.1)$$

An important and useful result is that if the population (data set) being considered is multivariate normally distributed, then the value of D^2 will follow a chi-squared distribution with p degrees of freedom (Manly, 1994). This means that a significantly large Mahalanobis value - one which is greater than the corresponding chi-squared distribution value at a particular significance level, for p (no. of explanatory variables) degrees of freedom – means that the corresponding observation is either (a) a genuine but unlikely record, or (b) a record containing some mistake. In other words a significantly large D^2 value indicates that the observation is an outlier from the rest of the population.

The performance of the Mahalanobis distance is highly dependent on the multivariate normality of the bulk of the data and Santos-Pereira and Pires (2002) described a method of removing this dependence. It is as follows:

a) Segment the n points cloud (n is the number of observations) into k smaller sub-clouds using a partitioning clustering method with the hope that each sub-cloud looks "more normal" than the original cloud.
b) Then apply a simultaneous multivariate outlier detection rule to each cluster by computing Mahalanobis-type distances from all observations to all the clusters. An observation is considered an outlier if it is an outlier for every cluster (greater than the chi-squared distribution value). All the observations in a cluster may also be considered outliers if the relative size of that cluster is small (less than 2p + 2).
c) Remove the observations detected in b, then repeat a and b until no more observations are detected.
d) The final decision on whether all the observations belonging to a given cluster (not previously removed, that is size greater than 2p + 1) are outliers based on a table of between clusters Mahalanobis-type distances.

The above method of outlier detection is only useful if an accurate calculation of the covariance matrix can be made. As the estimation of a covariance matrix can be inaccurate if there are a small number of degrees of freedom (number of observed operations) involved. Manly (1994) suggests this small number is less than 100 - which is often the

case in construction data. Additionally, construction practitioners would not have the time to utilise a complex statistical method. To solve these problems, an alternative method of identifying outliers has been devised in this thesis. This alternative method, is simpler than the Mahalanobis one, and involves replacing the complex Mahalanobis distance measurement with the simple Penrose distance (Penrose, 1953) of an observation from all other data points.

The Penrose distance is defined in equation 5.2 (notation can be found in appendix H).

$$P_{ij} = \sum_{k=1}^{p} \frac{(\mu_{ki} - \mu_{kj})^2}{pV_k} \tag{5.2}$$

for the distance between population *i* and population *j*.

The performance of the Penrose distance (like the Mahalanobis-type distance) is dependent on the normality of the bulk of the data, and although the data set in this study is normally distributed there are likely to be numerous data sets in construction that are not, thus a method which is equally applicable to both normal and non-normal data sets is presented in this thesis. To this end, it is proposed that steps (a) to (d) above be followed to identify outliers, but the Mahalanobis distance should be replaced with the simpler Penrose distance, in this simple method. A further simplification is that the Penrose distance does not follow the chi-squared distribution if the data is normal, and an observation will be declared an outlier if its Penrose distance from each cluster is apparently large compared with the range of Penrose distances of the remaining observations. To examine the adequacy of this simple method in identifying outliers, it was applied to a data set that contains known outliers and a comparison with the results from the Santos-Pereira and Pires (2002) method was made (see appendix B for full details). The simple outlier detection method was seen to outperform the Mahalanobis method in appendix B, and thus can be used with confidence on the RMC supply process data.

The simple method of identifying outliers (detailed above) shall be used to examine the non-normally distributed construction data (see 5.3.2). To re-affirm, the simple method involves:

- Undertaking a k-means cluster analysis of the data (an attempt to make the data more normal);
- Removing any strongly correlated variables;
- Calculating the Penrose distance between each data point and each cluster, and marking the extreme points as outliers. Similar to the grass data study (of appendix B) a number of k-values were used to attempt to build up a reliable picture of any outliers in the data sets. Figure 5.3 displays the effect of cluster analysis on the distribution of the process data, and it can be concluded that the data has been made more normally distributed using the cluster analysis.

The next stage – an assessment of the correlations between the variables - has been completed in this statistical analysis, but reported later in section 5.4.4. The results reported in 5.4.4 have been used in this outlier analysis. The final stage was to calculate the Penrose distance between all the data points and each cluster and examine these distance values - highlighting any extreme values as suspected outliers. The results of the examinations are shown in table 5.3.

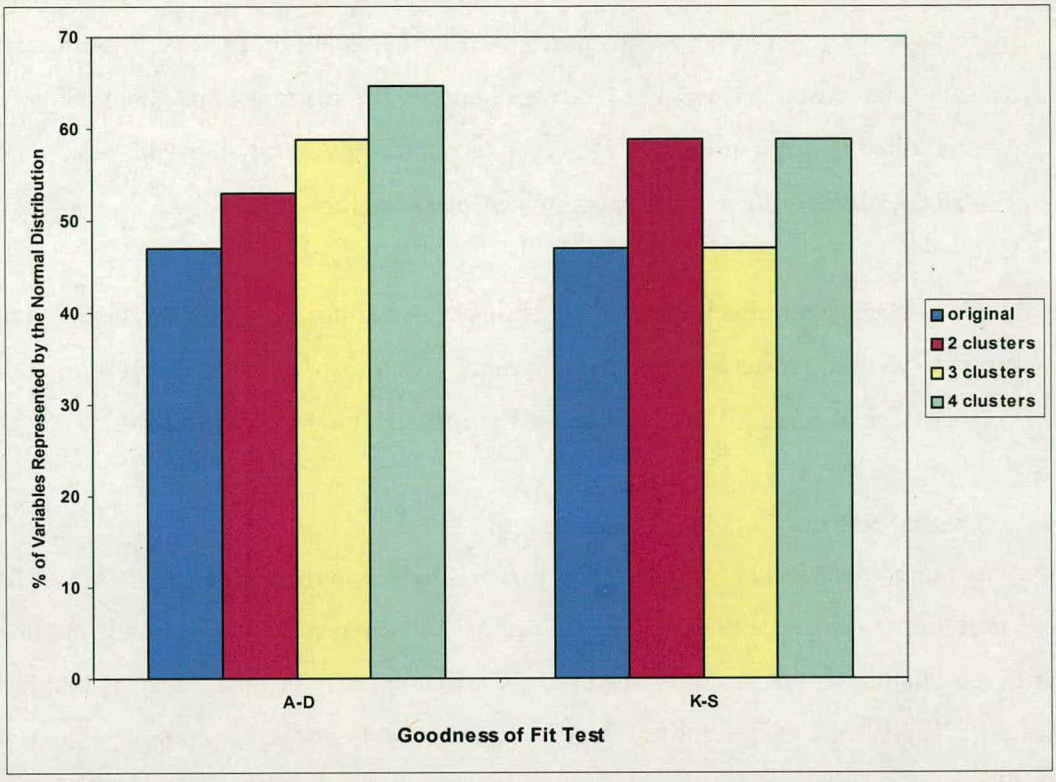


Figure 5.3: The effect of clustering on normality of the process data

Number of Clusters k		Cluster 1 Case	1 Penrose	2 Case	2 Penrose	3 Case	3 Penrose	4 Case	4 Penrose
2		41	4.7	33	8.0				
				18	5.9				
	Range of Non-Outliers		**3.7**		**3.0**				
3		33	6.5	83	25.7	41	7.7		
		205	5.7	18	24.0				
	Range of Non-Outliers		**4.1**		**22.3**		**7.0**		
4		205	43.4	83	43.8	205	52.2	12	28.9
								15	28.6
								83	28.1
	Range of Non-Outliers		**35.6**		**40.2**		**44.8**		**23.6**

Table 5.3: Data points with large Penrose distances from the clusters in the data

There were seven data points (cases) that were suspected of being outliers from one or more clusters. To illustrate the reasons that these cases were marked as suspected outliers we shall examine the results for cluster 1, with k = 4. In this instance, case 205 was found to have a Penrose distance of 43.4. If this data point were ignored, the remaining data set

would have a range of Penrose distances of 35.6. It was felt that the Penrose distance of case 205 from cluster 1 was large when compared with this range and therefore case 205 was suspected of being an outlier. However, as none of the seven observations are outliers from all the clusters, the process data can be declared outlier-free.

As it has been determined that no outliers exist in the process data, no further analysis using the techniques described above are required on the data, and in future non-robust statistics, such as LS regression, can be used with a high degree of confidence.

5.4.3 *Dealing with non-numerical variables*

The construction data sets contain variables that are non-numerical. There are two varieties of these non-numerical variables in the data: ordinal or ordered qualitative; and, nominal or unordered qualitative. An example of an ordinal variable is: x = income category, that is, low, medium or high. The date (month) of pour variable in the data set is an example of an ordinal variable. An example of a nominal variable is: x = marital status, that is, married, never married or previously married. The project, type of pour and concrete mix are examples of nominal variables.

To include non-numerical variables in most modelling methods, they must be given a score (numerical value). Choosing an adequate scoring method is difficult and the use of poor scores can, for example, increase dramatically the non-linearity of the relationship between the dependent and explanatory variables, hence making it more difficult for a model to provide an accurately estimate of a dependent variable (Myles, 1997). In an attempt to overcome these difficulties and identify an appropriate strategy for dealing with non-numerical variables, a LS regression analysis (using the step-wise method) was undertaken on the data when: the non-numerical variables were excluded from the data set; the non-numerical variables were scored using positive (>0) consecutive integers); the non-numerical variables were scored using a negative integer scale (<0). By examining the difference between the positive score regression and the negative score regression, the significance of the sign (positive or negative) may be ascertained. The results from the regression show (table 5.4) that including non-numerical variables with a positive scoring, afforded the greatest representation of the variability in the process (highest adjusted R^2 value). Thus, the data used subsequently includes formerly non-numerical data, now transformed into positive numeric form.

Concreting Process
Only For Cases with Missing Values Removed

Significance Level	Non-Numerical Variables Strategy		
%		Score Method	
	Variables Removed	Positive	Negative
5	78(5)	81(7)	80(8)
1	77(7)	80(9)	80(9)

Notes:
The values are the adjusted R^2 values from LS regression analysis

The values in brackets are the number of steps required to reach the final adjusted R^2 value.

Table 5.4: Non-numerical variables linear regression results

5.4.4 Dealing with correlated variables

Independent variables can be described as correlated if they have a linear relationship. Thus, a model will lose no knowledge of the system if one of these variables is not considered. Highly correlated variables can be identified easily by examining the Pearson, Kendall and Spearman correlation coefficients (Sprent, 1993) of each variable pairing. Although the values for each correlation coefficient type (i.e. Spearman, Kendall or Pearson) are not directly comparable, the trends are.

Variable	Project	Date (Month) of Pour	Travel Distance	Location of Pour	Type of Pour	Concrete Mix Type	Number of Truck-Mixers	Total Pour Volume	Truck-Mix Reference	Batch Volume	Average Volume/ Load	Time of Arrival at Batching Plant	Start Batching Time	Complete Batching Time	Batch Time	Time of Arrival on Site	Travel Time	Time Discharge Began	Time Discharge was Complete	Position Time	Discharge Time	Start Time of Washout	Completion Time of Washout	Washout Time	Pour Duration	Daily Maximum Temperature	Daily Minimum Temperature	Amount of Rain	Amount of Sunshine	Wind Speed	Grass Minimum Temperature
Project	1.0	0.0	0.4	0.3	0.3	0.0	0.3	0.5	0.4	0.4	0.4	0.5	0.4	0.4	0.0	0.4	0.2	0.2	0.3	0.4	0.5	0.0	0.4	0.2	0.6	0.5	0.4	0.2	0.3	0.4	0.0
Date (Month) of Pour		1.0	0.1	0.6	0.1	0.1	0.5	0.4	0.5	0.4	0.4	0.2	0.2	0.5	0.1	0.1	0.0	0.4	0.3	0.4	0.2	0.3	0.4	0.1	0.1	0.1	0.2	0.1	0.4	0.1	0.3
Travel Distance			1.0	0.4	0.5	0.6	0.4	0.5	0.1	0.4	0.3	0.1	0.4	0.4	0.5	0.5	0.7	0.5	0.1	0.5	0.3	0.4	0.5	0.2	0.2	0.1	0.1	0.4	0.2	0.1	0.5
Location of Pour				1.0	0.1	0.5	0.1	0.1	0.2	0.5	0.3	0.3	0.5	0.5	0.5	0.2	0.4	0.2	0.2	0.1	0.6	0.2	0.0	0.1	0.1	0.1	0.4	0.6	0.4	0.5	0.3
Type of Pour					1.0	0.5	0.4	0.1	0.5	0.0	0.2	0.1	0.4	0.3	0.5	0.2	0.4	0.4	0.5	0.4	0.5	0.4	0.2	0.3	0.0	0.1	0.4	0.2	0.4	0.2	0.4
Concrete Mix Type						1.0	0.1	0.5	0.3	0.1	0.3	0.4	0.4	0.2	0.1	0.0	0.3	0.2	0.6	0.4	0.4	0.3	0.4	0.6	0.2	0.0	0.2	0.2	0.3	0.2	0.4
Number of Truck-Mixers							1.0	0.6	0.0	0.4	0.6	0.3	0.4	0.2	0.1	0.5	0.1	0.1	0.0	0.2	0.1	0.1	0.5	0.2	0.4	0.1	0.3	0.4	0.2	0.0	0.1
Total Pour Volume								1.0	0.4	0.4	0.4	0.3	0.1	0.4	0.0	0.4	0.1	0.3	0.2	0.3	0.4	0.2	0.1	0.2	0.3	0.3	0.2	0.3	0.6	0.5	0.3
Truck-Mix Reference									1.0	0.1	0.4	0.3	0.5	0.5	0.0	0.6	0.2	0.2	0.2	0.4	0.1	0.2	0.3	0.0	0.5	0.2	0.2	0.4	0.1	0.5	0.1
Batch Volume										1.0	0.2	0.5	0.0	0.5	0.2	0.5	0.2	0.6	0.1	0.3	0.6	0.3	0.0	0.4	0.4	0.6	0.4	0.2	0.0	0.2	0.5
Average Volume/ Load											1.0	0.1	0.5	0.5	0.4	0.2	0.0	0.3	0.6	0.1	0.1	0.6	0.6	0.3	0.3	0.3	0.3	0.4	0.4	0.1	0.0
Time of Arrival at Batching Plant												1.0	0.9	0.7	0.2	0.5	0.2	0.2	0.5	0.1	0.2	0.0	0.3	0.1	0.6	0.2	0.3	0.3	0.6	0.5	0.4
Start Batching Time													1.0	0.9	0.6	0.4	0.5	0.3	0.2	0.0	0.1	0.1	0.5	0.4	0.2	0.6	0.3	0.0	0.5	0.1	0.3
Complete Batching Time														1.0	0.4	0.4	0.4	0.3	0.4	0.2	0.1	0.0	0.1	0.1	0.0	0.1	0.6	0.1	0.1	0.0	0.4
Batch Time															1.0	0.5	0.2	0.5	0.4	0.2	0.5	0.1	0.1	0.5	0.6	0.1	0.2	0.4	0.3	0.3	0.1
Time of Arrival on Site																1.0	0.1	0.9	0.7	0.4	0.5	0.5	0.6	0.0	0.2	0.3	0.5	0.3	0.6	0.3	0.2
Travel Time																	1.0	0.5	0.5	0.1	0.1	0.3	0.1	0.3	0.6	0.5	0.6	0.6	0.5	0.5	0.3
Time Discharge Began																		1.0	0.1	0.1	0.3	0.4	0.3	0.4	0.3	0.6	0.5	0.6	0.5	0.2	0.3
Time Discharge was Complete																			1.0	0.3	0.5	0.3	0.5	0.3	0.4	0.2	0.4	0.1	0.5	0.0	0.3
Position Time																				1.0	0.6	0.3	0.4	0.0	0.1	0.4	0.1	0.4	0.6	0.1	0.0
Discharge Time																					1.0	0.1	0.4	0.3	0.3	0.4	0.5	0.2	0.1	0.1	0.5
Start Time of Washout																						1.0	0.6	0.3	0.2	0.4	0.5	0.4	0.1	0.5	0.4
Completion Time of Washout																							1.0	0.4	0.1	0.6	0.4	0.0	0.5	0.5	0.1
Washout Time																								1.0	0.3	0.5	0.3	0.2	0.2	0.1	0.6
Pour Duration																									1.0	0.2	0.4	0.2	0.0	0.3	0.3
Daily Maximum Temperature																										1.0	0.5	0.4	0.6	0.2	0.3
Daily Minimum Temperature																											1.0	0.4	0.2	0.2	0.9
Amount of Rain																												1.0	0.3	0.4	0.6
Amount of Sunshine																													1.0	0.2	0.6
Wind Speed																														1.0	0.2
Grass Minimum Temperature																															1.0

Table 5.5: Pearson correlation coefficients of the RMC supply process variables

Table 5.5 shows the Pearson correlation coefficients for the independent variables of the RMC supply process. It was found that the Pearson, Kendall and Spearman correlation coefficients agree that minimum temperature is highly correlated with minimum grass temperature. This conclusion is supported by the linear relationship shown in the scatter plot in figure 5.4, and therefore a decision has been made to remove the grass minimum temperature variable from the data set. The other process variables that displayed linear relationships were the time of arrival at batching plant and the start batching time – the time of arrival variable has also been removed from the data set. There are now 24 independent variables remaining in the data set.

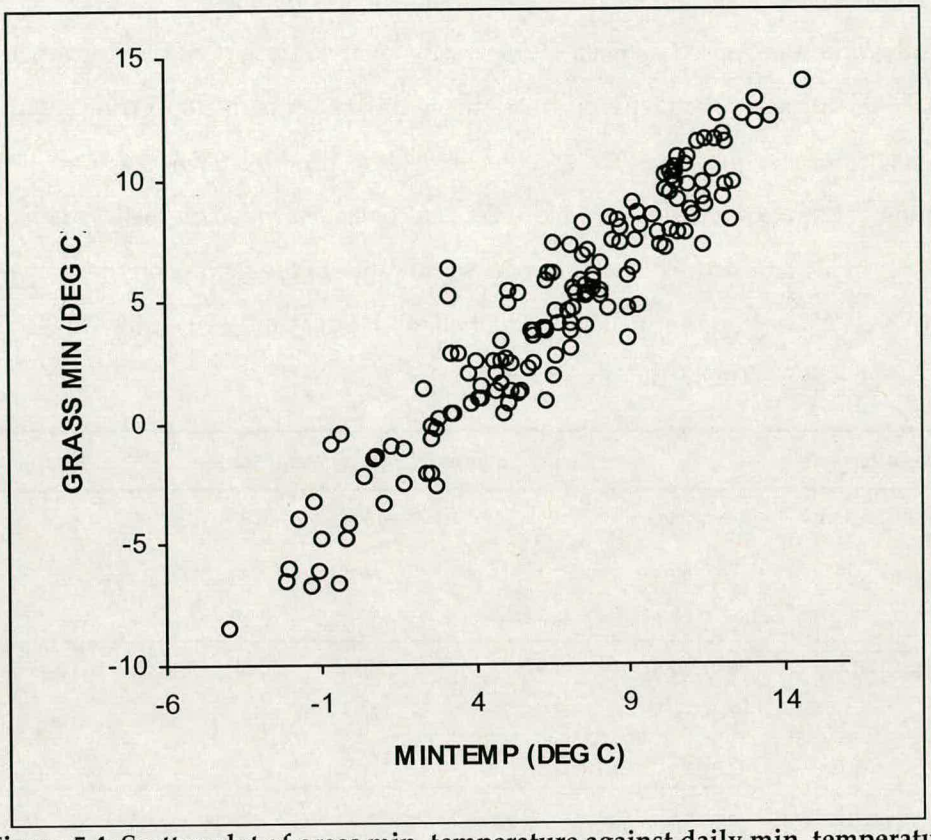


Figure 5.4: Scatter plot of grass min. temperature against daily min. temperature

5.5 Examining the nature of the relationship between the dependent and explanatory variables of the process

There are two questions that must be answered regarding the characteristics of the process variables. They are:

- Is there any indication of a systematic relationship between the dependent and explanatory variables?
- Is a linear model capable of providing accurate and reliable estimates of the process?

With respect to these questions, it is critical that the answer to the first is "yes". Otherwise, there is no modelling technique which will be able to describe the process adequately. The second question emphasised the principle of parsimony, which should underlie the modelling of any process. Therefore it is important to ascertain whether or not a simple, linear model of the RMC supply process is adequate, before anything more complex is to be considered.

5.5.1 Identification of a systematic relationship between the variables

To answer the questions posed above a LS regression analysis was performed for each of the output variables (process productivity, truck-mixer utilisation and pump utilisation) of the process. This means an equation of the type y = mx + c was produced, where y is an output, m is a weighting, x is an explanatory variable and c is a constant defined by the LS mathematical function. To provide a measure of the statistical significance of the regression fit, an analysis of variance (ANOVA) was performed (Mendenhall and Sincich, 1995) using the statistical software package, SPSS. The results are shown in table 5.6.

Dependent Variable		Sum of Squares	df	Mean Square	F	Significance
Pump Utilisation	Regression	4934	15	329	3	0
	Residual	12617	126	100		
	Total	17552	141			
Truck Utilisation	Regression	19923	15	1328	6	0
	Residual	26723	126	212		
	Total	46646	141			
Process Productivity	Regression	3603	15	240	42	0
	Residual	726	126	6		
	Total	4329	141			

Table 5.6: ANOVA table for linear regression analysis of the RMC supply process

The critical value of the F-distribution (Lindley and Scott, 1984) at a confidence level of 99% is 2.3 for the RMC supply process. This value is based upon the number of data points available (232 in this case). It was observed that for each of the dependent variables, the F-value (test statistic) exceeded the critical value. This means that it can be said with at least a 99% confidence that an approximate linear relationship exists between the dependent variable and one or more explanatory variables. A caveat to this is that if the relationship between the dependent and all of the explanatory variables is non-linear, as is reported in previous publications (e.g. Smith, 1998), the confidence level of 99% is not precise. Nonetheless, the ANOVA results provide a strong indication of the existence of a systematic relationship between the dependent and explanatory process variables.

5.5.2 Quality of the linear fit

Having confirmed the existence of a systematic relationship between the process variables, it is important to satisfy the second question of 5.5 and determine if a linear model will adequately produce estimates of the dependent variables. Again the results from the linear LS regression analysis were used in this investigation.

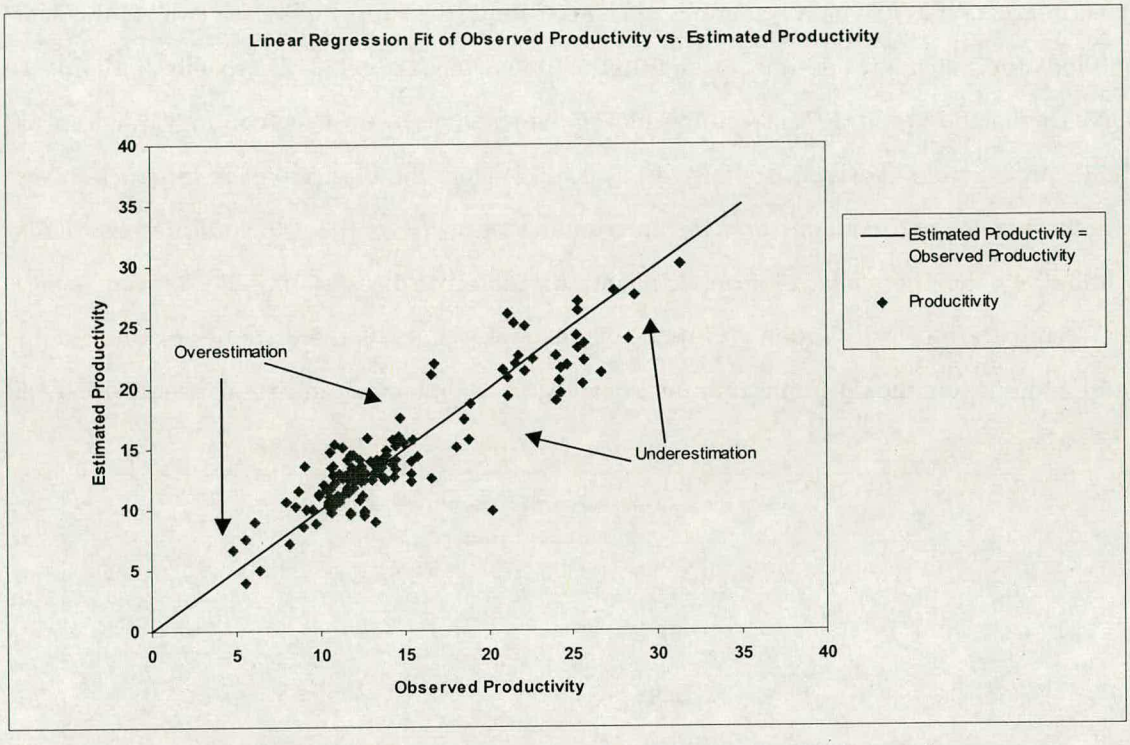


Figure 5.5: LS regression estimation of productivity of the RMC supply process

The plot concerning process productivity (see figure 5.5) demonstrates a partial linear trend in the data. However, for lower productivities the fit tends to overestimate, and for higher productivities the fit underestimates. Thus, it is unclear whether a linear model would be entirely suitable for modelling the productivity of the RMC supply process, and further investigation of any non-linear traits is warranted (appendix C).

The other dependent variables, the truck-mixer and pump utilisation rates, were found to have suspected non-linear relationships with the independent variables. Further analysis is required concerning these variables, as the inaccuracy of the linear fit may be due to important explanatory variables not being included in the data set. If that were the case, non-linear

modelling will not be effective, and the only alternative is to collect new data containing more variables.

5.5.3 *Confirming a non-linear relationship for the other dependent process variables*

To confirm the existence of non-linear relationships between the other dependent variables and the explanatory variables, a technique standard to identifying the need for using a non-linear regression model has been applied. This technique is to apply a linear model to the data and look for systematic non-linear trends in the fit residuals (Fox, 1984; Anscombe and Tukey, 1963; Daniel and Wood, 1971). A scatter plot of the residuals against each of the variables was made, an example of which is shown in figure 5.6. Here the scatter plot is for truck-mixer utilisation rates and contains an 11-point running mean (Fox, 1984) to indicate the trends within the data. The scatter plot provides an indication that there is a non-linear relationship between truck-mixer utilisation and the explanatory variables as there are no less than eight local minima on the 11-point running mean, and a global minimum at around 67-72% utilisation.

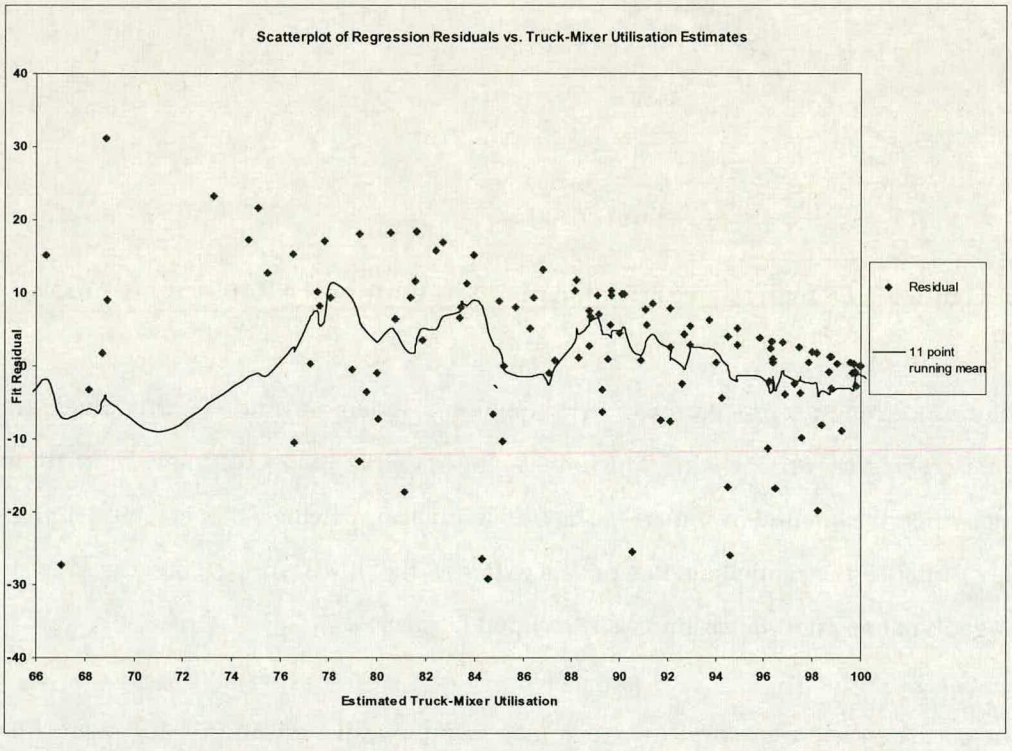


Figure 5.6: Scatter plot of fit residuals against estimated truck-mixer utilisation

To substantiate this conclusion another scatter graph was plotted (figure 5.7), but this time with three statistical measures included as this reportedly provides enough information to

summarise a scatter plot (Cleveland and Kleiner, 1975). The three measures used were: the mid-mean (Tukey, 1977), the mean of the utilisations within the inter-quartile range; and, the upper and lower semi mid-means (Tukey, 1977), the mean of the utilisations above the upper quartile and below lower quartile, respectively. It is sometimes desirable to further smooth these statistical measures (Cleveland and Kleiner, 1975) and in this case, due to erratic behaviour displayed by the mean in figure 5.7 (after smoothing), this was carried out using an 11-point running average. The results indicate that there is a non-linear relationship as the mid-mean reaches a number of local minima points, and the variance in the data varies from being large between utilisations of 79 and 89%, to significantly less outside this range. This examination was undertaken for the final dependent variable, pump utilisation and a similar result was discovered.

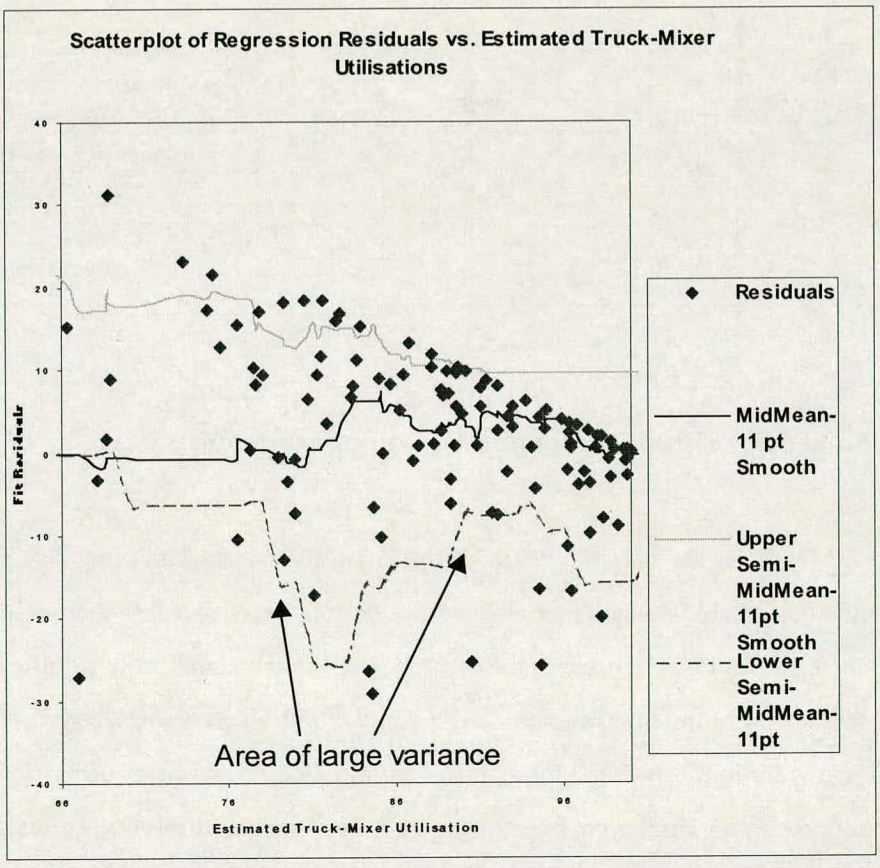


Figure 5.7: Scatter plot of fit residuals against estimated truck-mixer utilisation

5.5.4 A further examination of the non-linear relationship between the process variables

The results from the previous section indicated that there were non-linear relationships between the process productivity, truck-mixer and pump utilisations and the explanatory

variables. The method used to identify these trends was simple, but can prove insensitive to many types of non-linearity, therefore a more robust method that can provide stronger evidence of non-linear relationships must be used (Myles, 1997). One robust method is to scatter plot the residuals against the explanatory variables and look at the trends in the data. Several of the scatter plots display non-linear trends in the residuals, which can be described as statistically significant. Figure 5.8 shows such a plot for the linear fit of truck-mixer utilisation.

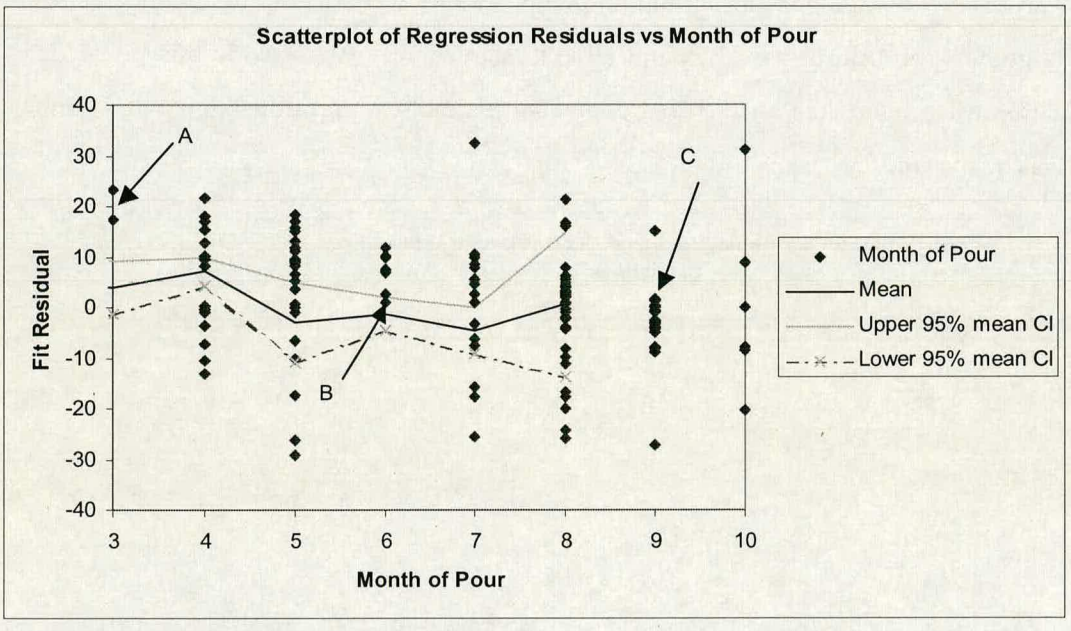


Figure 5.8: Scatter plot of truck-mixer utilisation regression residuals vs. month of pour

The residual bins (explanatory variable value with a large number of data points) A, B and C have means which alternate between being significantly greater than or less than zero. In the plot for truck-mixer utilisation (figure 5.8) the captions highlight a statistically significant non-linear trend - which is bump (between A and B) and bowl shaped (between B and C) - providing strong evidence for the need for a non-linear model of truck-mixer utilisation rate in the RMC supply process. A similar conclusion can be drawn for the pump utilisation rate (not shown here) and the productivity of the RMC supply process (see appendix C).

5.6 Choosing model variables using a modelling framework

A framework to be followed in the development of any models in this PhD project was devised and is shown in figure 5.9. The aim of this modelling framework is to allow models to be developed of the RMC supply process that can provide accurate estimates of the process performance at the time when an order for RMC is placed with a supplier, which is roughly 4 days prior to construction. This time shall be known as the 'order time' for the remainder of this thesis. The above aim is achieved by using information relating to a specific RMC pour that is realistically known by a construction practitioner at the order time.

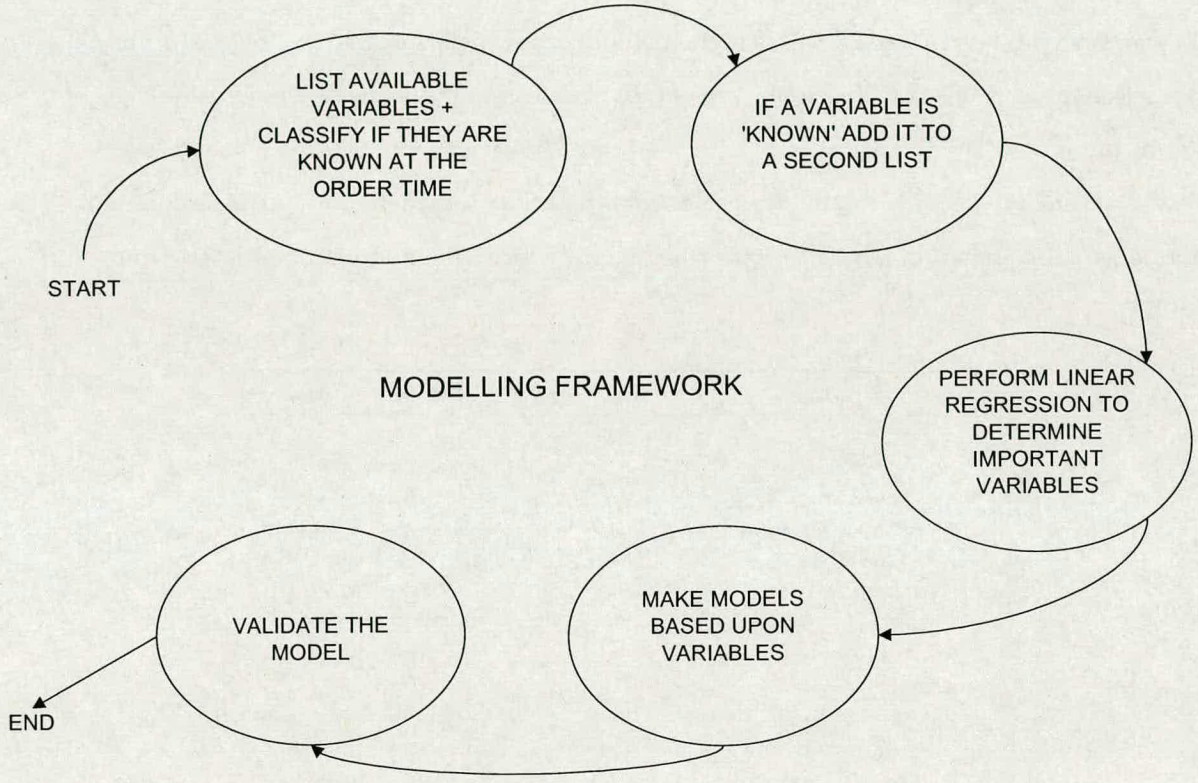


Figure 5.9: Modelling framework that was followed in model development

The first three stages of the modelling framework shall be discussed in this chapter; model development and validation shall be focussed upon in the next chapter.

5.6.1 *Modelling framework stages one and two: classifying the 'known' variables and listing them for further consideration*

The initial stages of the framework involve classifying the available variables according to their level of determinability, by a construction practitioner at the order time. This process was performed under consultation with a professional construction manager, Mr. Doug Forbes, from MJ Gleeson (Forbes, 2004). Thereafter, all highly determinable variables are to be considered further, with the remainder being discarded.

The available variables of the RMC supply process were listed earlier in table 5.1. Here, each of these variables has been assessed with a high, medium or low determinability. Obviously, this is a subjective decision, and that is the reason that professional help was acquired for this stage of the thesis. The results of this assessment are shown in table 5.7. If a variable was assessed as highly determinable, it was selected for further consideration and placed in a separate list (also shown in table 5.7). This assessment reduced the number of variables from 24 to 12.

Variable	Determinability	Remaining Variables
Project	High	Project
Date (Month) of Pour	High	Date (Month) of Pour
Travel Distance	High	Travel Distance
Location of Pour on Structure	High	Location of Pour on Structure
Type of Pour	High	Type of Pour
Concrete Mix Type	High	Concrete Mix Type
Number of Truck-Mixers	High	Number of Truck-Mixers
Total Pour Volume	High	Total Pour Volume
Truck-Mixer Reference Number	Low	Batch Volume
Batch Volume	High	Average Volume / Load
Average Volume / Load	High	Daily Maximum Temperature
Time of Arrival at Batching Plant	Medium	Daily Minimum Temperature
Start Batching Time	Medium	
Complete Batching Time	Medium	
Time Discharge Began	Medium	
Time Discharge was Complete	Medium	
Start Time of Washout	Medium	
Completion Time of Washout	Medium	
Pour Duration	Medium	
Daily Maximum Temperature	High	
Daily Minimum Temperature	High	
Amount of Rain	Medium	
Amount of Sunshine	Medium	
Wind Speed	Medium	

Table 5.7: Steps one and two of the modelling framework applied to the process

5.6.2 *Modelling framework stage three: LS regression analysis of the remaining variables*

The list of 'known' independent variables in table 5.7 are now used in a linear regression analysis (as the x component) against the three important outputs of the RMC supply process, process productivity, pump utilisation and truck-mixer utilisation (y components). The result of these analyses is that three sets of important variables (the ones remaining after regression) are defined (one for each output). These lists of important variables must be combined into one list. This means that only one model will be developed and not three. To facilitate this reduction, the variables which were selected as important for all outputs are to be defined as important in the overall RMC supply process, and thus shall be used in modelling. For example, the month of pour was found to be important in the LS regression analysis of the process productivity, pump utilisation and truck-mixer utilisation variables – and was consequently selected as a variable for use in modelling. On the other hand, the location of a pour on the structure was found to be important to the truck-mixer utilisation only. Thus, it was ignored in the next stage of the framework. The final set of variables is shown in table 5.8 below. The number of variables has been reduced again, from 12 to 6.

Variables to be Used in Modelling
Date (Month) of Pour
Travel Distance
Type of Pour
Number of Truck-Mixers
Total Pour Volume
Daily Maximum Temperature

Table 5.8: The variables to be used in model development

The ratio of the amount of data to the number of variables has increased to 232:6 and this should improve the likelihood of modelling the RMC supply process successfully.

5.7 Chapter summary and conclusions

In this chapter a statistical analysis was performed on the collected RMC process data. The findings and conclusions from this chapter are:

- There are 26 explanatory variables which were suspected of holding a relationship with the dependent process variables.
- A majority of the process variables were found to be normally distributed, meaning that standard statistical methods were applicable to the data.
- The data set contained missing values and it was determined that these values should be estimated using a method known as 'linear interpolation', resulting in a larger set of data than originally collected.
- A new method of detecting outliers in construction data sets was devised and presented. This method found that there were no outliers in the collected data set. This result means that robust methods were not required in this statistical analysis.
- It was found that converting non-numerical variables using the outlined positive-integer scoring method was the best way to deal with non-numerical variables.
- Two sets of correlated variables were identified, and thus two variables (one from each set) were removed from the process data set. 24 explanatory variables remained in the data set.
- The explanatory variables of the RMC process were found to hold a non-linear relationship with the process productivity, pump utilisation and truck-mixer utilisation. Thus, to model any of these dependent variables requires a non-linear modelling method, such as simulation.
- To develop a set of variables which shall be used in modelling, a framework was devised. This framework places great importance on only including the variables which would be known by construction practitioners at the time of ordering RMC, which is usually 4 days prior to construction (representing RMC supplier's lead-time). Using this framework a set of six variables was produced for modelling.

CHAPTER 6

SIMULATION MODEL DEVELOPMENT AND VALIDATION

Having established the presence of non-linear relationships between the variables of the RMC process, a modelling technique needs to be established and developed. Discrete-event simulation is an established technique for investigating such problems, and this chapter explores its usefulness.

6.1 Introduction

In the preceding chapters, the UK construction industry was thought to be having difficulty fulfilling the orders placed upon it in a timely and cost-effective manner. This is in part due to the poor performances that have been and still are witnessed in the cyclical, stochastic construction processes that are prevalent in many construction projects. This project has focussed upon modelling construction at the process level, and initially one process in particular – the RMC supply process.

The current method of intuitively estimating the number of resources required in the RMC supply process is thought to be the root cause of poor performance; often processes are under – or – over resourced. This project has focused upon developing a computer-based, discrete-event simulation model to remove the need for these intuitive estimates to be made. Such models have been developed in the past (see section 6.3), and although many of them have proven capable of providing fairly accurate estimates of cyclical, stochastic construction processes, they have generally lacked the capability of being easily implemented in practice. Thus, this project intends to develop a model that can accurately estimate the RMC supply process in a practical manner.

Chapter 6 is structured as follows:

- **Section Two.** What is discrete-event simulation?
- **Section Three.** Past use of discrete-event simulation in construction.
- **Section Four.** A discrete-event simulation model of the RMC supply process is presented. The development of the model is discussed, as is the evolution of the model over the project duration.
- **Section Five.** Fitting probability distributions to the times of interest of the RMC supply process.
- **Section Six**. The discrete-event simulation model is validated to assess its capability of recreating the RMC supply process.
- **Section Seven**. Conclusions from this chapter are drawn.

6.2 What is discrete-event simulation?

Simulation is a modelling procedure that can be used to recreate a stochastic process. It involves the design of a model of the process and the performance of experiments on that model to learn about particular aspects of the process (Martinez, 1996; Pritsker et al., 1997). This could entail estimating the behaviour of a process under certain conditions, then re-estimating this behaviour if any changes in the conditions were to occur. For example, the productivity (behaviour) of the RMC supply process can be estimated for a pour involving 4 truck mixers (condition). If the number of trucks involved in a pour were to increase to 6 truck mixers (new condition), then a simulation model could estimate the productivity of the RMC supply process given this new condition, allowing the impact on behaviour of changing the condition to be calculated.

Discrete-event simulation concerns modelling a process over time by a representation in which the state of that process (i.e. server is idle, server is busy) changes instantaneously at separate points in time. These points mark the occurrence of an event, where an event is defined as an instantaneous occurrence that may change the state of the system (e.g. an arrival/departure of a customer at/from the server). Time in a discrete-event simulation model is recorded in a variable known as the simulation clock, which holds no relationship with the time required to run a simulation model. The simulation clock is advanced with the occurrence of each new event until some stopping condition is fulfilled, at which point a report of the relevant statistics is generated. A more detailed description of the components of a discrete-event simulation model and the relationships between these components is disclosed in section 6.2.1.

6.2.1 Components of a discrete-event simulation model

There are a number of components which are generic to every discrete-event simulation model (Law and Kelton, 2000). They are:

- **Process state**: The collection of state variables necessary to describe the process at a particular time.
- **Simulation clock**: A variable giving the current value of simulated time.
- **Initialisation routine**: The simulation clock is set to zero, the process state is initialised, the times for the next event of each type are calculated, the event list is initialised with these calculated times.

- **Event list**: A list containing the next time when each type of event will occur.
- **Statistical counters**: Variables used for storing statistical information about process performance.
- **Timing routine**: A subprogram that determines the next event from the event list and then advances the simulation clock to the time when that event is to occur.
- **Event routine**: A subprogram that updates the process state when a particular type of event occurs (there is one event routine for each event type).
- **Library routines**: A set of subprograms used to generate random observations from probability distributions that were determined as part of the simulation model.
- **Stopping condition**: A pre-defined criterion which when met results in the termination of the simulation model.
- **Report generator**: A subprogram that computes estimates (from the statistical counters) of the desired measures of performance and produces a report when the simulation ends.
- **Main program**: A subprogram that invokes the timing routines to determine the next event and then transfers control to the corresponding event routine to update the process state appropriately. The main program may also check for termination and invoke the report generator when the simulation is over.

The logical relationships between these components are shown in figure 6.1. These relationships can be summarised as follows:

- Simulation commences at time 0 with the main program invoking the initialisation routine, where the simulation clock is set to zero, the process state (i.e. server is made idle, customers are made available) and the statistical counters are initialised, and the event list is initialised.
- After control has been returned to the main program it invokes the timing routine to determine which type of event is imminent.
- If an event of type i (e.g. arrival of a customer at a server) is next to occur, the simulation clock is advanced to the time that event type i will occur (taken from the event list) and control is returned to the main program.
- Then the main program invokes an event routine i (such as those shown in figures 6.2 and 6.3), where typically three types of activity occur: (1) the process state is updated to account for the fact that an event type i has occurred; (2) information about process

performance is gathered by updating the statistical counters; (3) the times of occurrence of future events are generated and this information is added to the event list. Often it is necessary to generate random observations from probability distributions in order to determine the times of these future events; here, such generated observations will be referred to as a random variate.

- After all processing has been completed either in the event routine or in the main program, a check is typically made to determine if the simulation should now be terminated. This check is performed by assessing if the stopping condition has been fulfilled.
- If the stopping condition has been fulfilled, then the simulation is terminated, the report generator is invoked from the main program to compute estimates (from the statistical counters) of the desired measures of performance and to produce a report.
- If the stopping condition has not been met, control is passed back to the main program and the cycle between the main program- timing routine- main program- event routine- termination check is repeated until the stopping condition is satisfied.

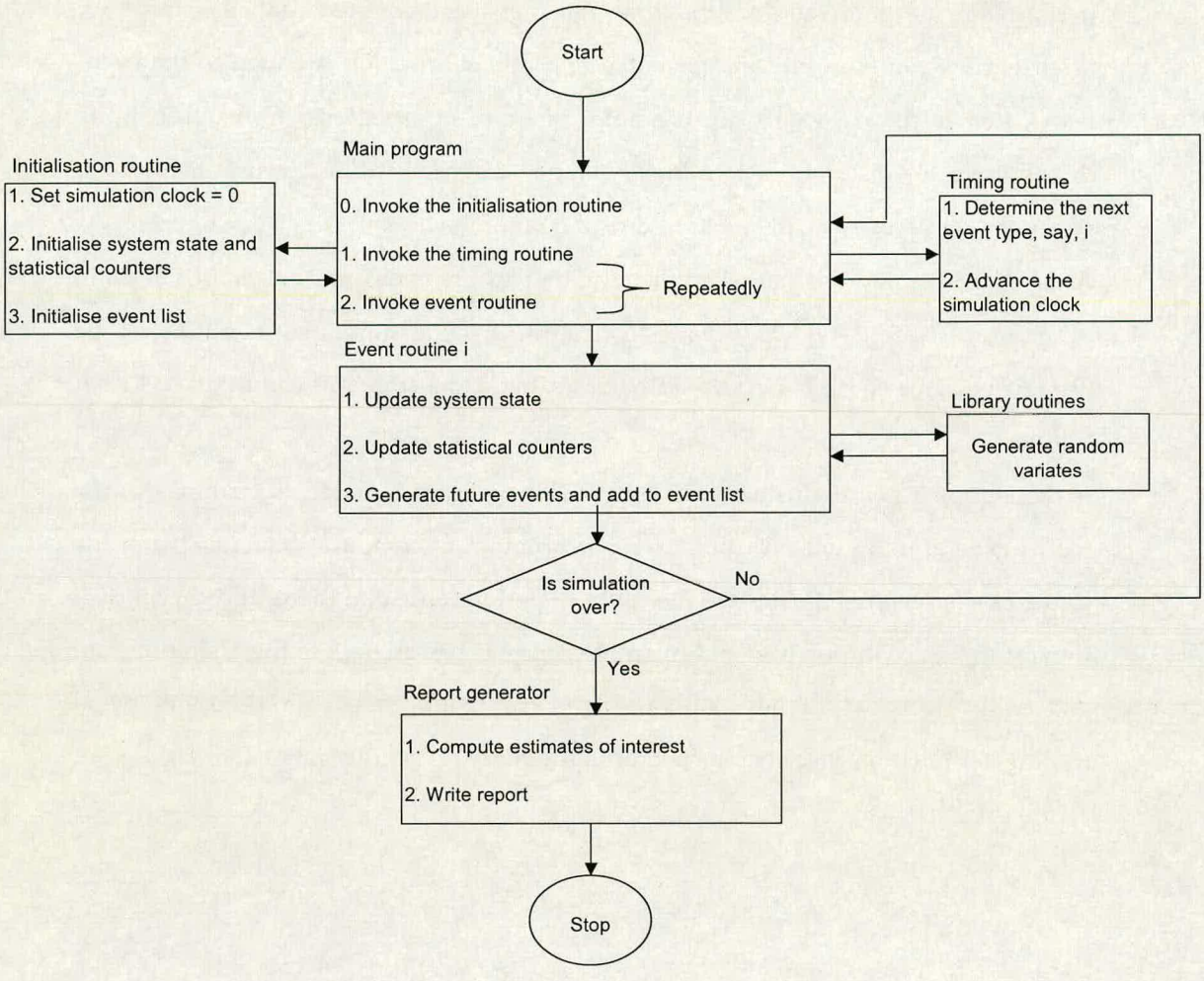


Figure 6.1: Relationship between discrete-event simulation model components (after Law and Kelton, 2000)

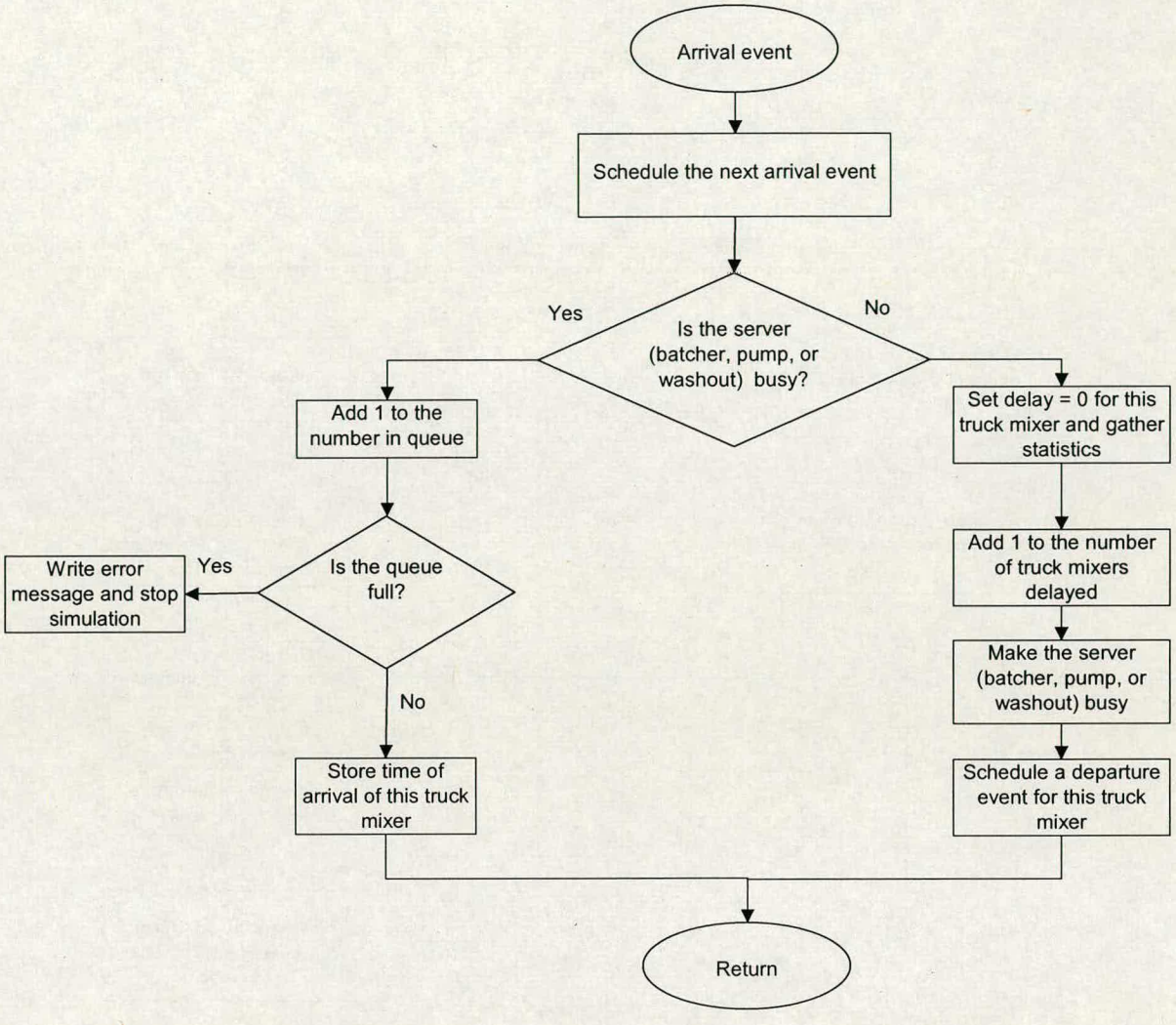


Figure 6.2: An arrival event routine (after Law and Kelton, 2000)

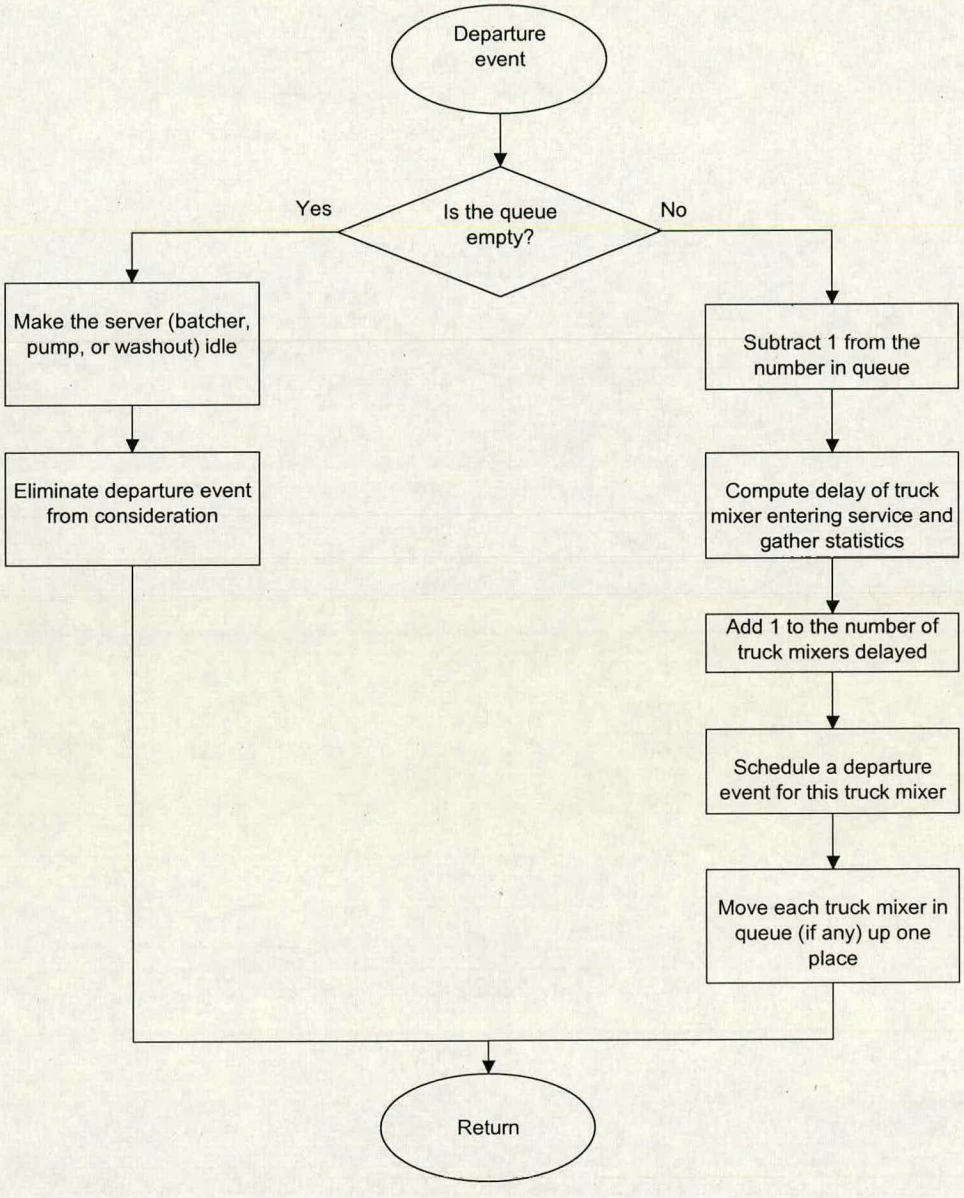


Figure 6.3: A departure event routine (after Law and Kelton, 2000)

6.3 Discrete-event simulation use in construction

Discrete-event simulation has been used extensively in construction research over the past three decades (Martinez and Ioannou, 1999). To demonstrate this fact a literature survey and review carried out among five international peer-reviewed journals (Journal of Construction Engineering and Management; Journal of Computing in Civil Engineering; Civil Engineering and Environmental Systems; Engineering Construction and Architectural Management; and

Computer-Aided Civil and Infrastructure Engineering) of the years that this PhD project has spanned (2002 to 2005). Over this period 30 research papers were published relating to the simulation of construction. These articles have been categorised into project modelling, process modelling and construction design. The categories are shown in table 6.1. Clearly a significant amount of research has been undertaken into the simulation of construction processes, but the method has not been widely used in the construction industry beyond large contractors who can afford to employ simulation specialists. This thesis is concerned with construction process modelling partly for this reason.

The use of discrete-event simulation in construction process modelling can be broadly categorised according to the computing method used to develop a model. There are three main computing methods that have been used in construction-related discrete-event simulation modelling (Smith, 1998):

- A simulation language.
- A general-purpose programming language.
- A spreadsheet package, such as *Microsoft Excel*.

Simulation Type	References
Project Modelling	Zhang et al. (2002a) Isidore and Back (2002) Lu (2002) Ammar and Mohieldin (2002)
Process Modelling	Hong et al. (2002) Leu and Hung (2002) Anson et al. (2002) Shi (2002) Marzouk and Moselhi (2002) Zhang et al. (2002b) Hajjar and Abourizk (2002) Kim and Gibson Jr. (2003) Dawood and Marasini (2003) Dawood et al. (2003) Cheng and Feng (2003) Nassar et al. (2003) Yang et al. (2003) Lu (2003) Lu et al. (2003) Hegazy and Kassab (2003) Zhang et al. (2003) Zhang et al. (2004) Feng et al. (2004) Huang et al. (2004) Cheng et al. (2005)
Construction Design	Clayton et al. (2002) Ozel and Kohler (2004) Malkawi (2004) Ren et al. (2004) Reichard and Papamichael (2005)

Table 6.1: Literature survey and review of 5 journals from 2002 to 2005

6.3.1 Simulation languages

A simulation language is a computer package that allows the simulation of a specific class of systems or processes with little or no programming. Simulation languages are often for a general purpose e.g. ARENA or EXTEND, but can be domain specific (Law and Kelton, 2000). Either way there are two main approaches adopted in the development of a simulation language (Naji and Najafi, 1996):

- network-based, where basic modelling elements (resources, etc.) are assembled to construct a network that represents a system or process under study allowing simulation of the network to be performed;

- graphical-based, where the geometry surrounding the process and the interactions between the process resources are included in the model.

Since its conception, construction simulation has been advancing with computer technology (Shi, 1999). Here follows a brief review of some of the major network-based construction simulation languages:

CYCLONE

In the early 1970s Daniel Halpin developed one of the first network-based construction simulation languages, *CYCLONE*, at the University of Illinois, Urbana (Halpin, 1973; 1977). *CYCLONE* manages construction simulation through simple rules defined within the simulation language, making it easy to use as both an analysis and communication tool. It may be easy to set up an appropriate model of a construction process with *CYCLONE*, but the language does not allow modelling to the level of detail required to make decisions, as the state of the system is hidden and resources cannot possess individual characteristics (Martinez, 1996).

The above drawbacks to *CYCLONE* notwithstanding, numerous construction processes have been modelled (using *CYCLONE*), including concrete batching (Woods and Harris, 1980; Lluch and Halpin, 1982) and tunnelling (Touran and Asai, 1987). Furthermore, there have been at least three implementations of the simulation language itself in different forms: *INSIGHT* (Paulson, 1978; Kalk, 1980; Paulson et al., 1987), *UM-CYCLONE* (Ioannou, 1989) and *Micro-CYCLONE* (Halpin, 1990).

RESQUE

RESQUE (Chang, 1986) is a simulation language that was developed to significantly improve the modelling capabilities and hence accuracy of *CYCLONE*. *RESQUE* permits resources to possess individual characteristics, resulting in a more realistic model of a process. Furthermore, the current state of the system (which resources are idle and which are busy) is presented to the user and this aids model verification (confirming the process has been modelled correctly) as the process is more transparent.

The capabilities of *RESQUE* were extended by Odeh (1992) in the development of *CIPROS*. This is a simulation language capable of planning both at a project and a process level.

STROBOSCOPE

STROBOSCOPE was developed by Martinez (1996) during his doctoral work at the University of Michigan. The aim of this simulation language was to improve upon the preceding languages by allowing a model to be simulated to a high degree of detail – resources possess individual characteristics and the state of the simulation model is transparent. A further aim was to broaden the use of simulation and to a degree this has been successful, as *STROBOSCOPE* is being used in the classrooms of many American, Asian and European Universities.

Opdenbosch and Baker (1994) criticised network-based simulation claiming it has failed to describe the three factors that influence many construction operations: the construction site; the geometry of the structure; and, the dynamics of construction equipment. For these reasons graphical simulation languages are becoming more prevalent. This type of simulation language involves drawing interactively on-screen the process to be modelled and the elements (resources, etc.) contained within the process. It often involves a graphical display of the process during simulation to allow a construction practitioner to visualise the situation. One significant drawback to graphical-based construction simulation is that the time it takes to produce a result may be prohibitive in a practical situation.

Notable graphical construction simulation languages include:

COOPS

COOPS was developed by Liu (Liu, 1991; Liu and Ioannou, 1992) as an object-orientated extension to *CYCLONE*. A graphical user interface (GUI) allows a user to select the objects (e.g. truck mixers) to be used in the model. A model is developed and the simulation results are animated.

DISCO

DISCO was developed by Huang and Halpin (Huang and Halpin, 1994; Huang et al., 1994). It is both a pre- and post- processor to *Micro-CYCLONE*. The pre-processor consists of a GUI

similar to the one developed for *COOPS*. The post-processor creates an animation of any simulation experiments performed.

VIS

Visual Interactive Simulation (VIS) is a language adopted from the field of virtual reality (Naji and Najafi, 1996). *VIS* uses virtual reality technology to model, design and simulate construction processes. An implementation of the *VIS* language was made fairly recently at the Georgia Institute of Technology known as *Interactive Visualiser++ (IV++)* by Opdenbosch and Baker (1994). The *VIS* language has not been widely used because of the expensive operating system that is usually required to run virtual reality programs (Naji and Najfi, 1996; Naji, 2000).

6.3.2 General-purpose programming languages

A general-purpose programming language such as *C, C++, Fortran, Visual Basic (VB), Matlab or Java* may be used to develop simulation models of construction processes. Although this enterprise is time-consuming, as models must be built from scratch, this effort is rewarded, as a developer will know the workings of the model in great detail. One significant drawback is that a model developed using this type of programming language is sometimes not generic enough to model different processes (Martinez, 1996). This is counteracted by the fact that greater flexibility is afforded to the process being modelled, as more details can be included in the model (Pritsker, 1998). Another advantage of software generated using a general-purpose programming language is that it is usually more economical than specific simulation language packages (Law and Kelton, 2000).

One of the first construction simulation models developed with a general-purpose computer programming language was created to aid equipment selection by Teicholz (1963). Others have been developed over the years, including: a model to evaluate the uncertainty in time and cost of underground construction (Moavenzadeh and Markow, 1976); a model to estimate project duration (Carr, 1979); a model to assess alternative resource allocation strategies (Morua, 1986); and a model to simulate underground geological conditions (Ioannou, 1984). More recently, *Trucksim* a C-based model of earthmoving operations has been developed at the University of Edinburgh (Smith, 1995; Smith et al., 1995).

6.3.3 A spreadsheet package

A spreadsheet package, such as *Microsoft Excel* can be used to create a simulation model of construction processes. Modelling on a spreadsheet allows the developer and a user to more fully understand (than when using the methods detailed above) the steps involved in both the process and the model of that process, thus fundamental mistakes can be avoided. One drawback to using a spreadsheet package is that the model is neither particularly user-friendly nor flexible.

Simulation models of the concrete supply and use processes have been created using *Microsoft Excel* in the past by Smith (Smith, 1998; Smith, 1999).

6.4 Simulation model development

In this section the development of a discrete-event simulation model of the RMC supply process is detailed. Before these details are provided it is worthwhile to recap on how the RMC supply process is to be viewed in such a model. Adapting the process shown in figure 3.2 to include the times of interest of the process can provide this view. This new view of the process is shown in figure 6.4.

To summarise figure 6.4, the RMC supply process is thought has consisting of three FIFO single-server queues: one at the batching plant; one at the concrete pump; and one at the washout area. The *times of interest* are:

- The time to batch a load of RMC into a truck mixer (or batch time);
- The time to travel to the project site (travel to site time);
- The time to position the truck mixer at the concrete pump (position time);
- The time to discharge the load from the truck mixer into the concrete pump (discharge time);
- The time to clean out the residue from the nearly empty truck mixer (washout time);
- The time to travel from the project site back to the batching plant (travel from site time).

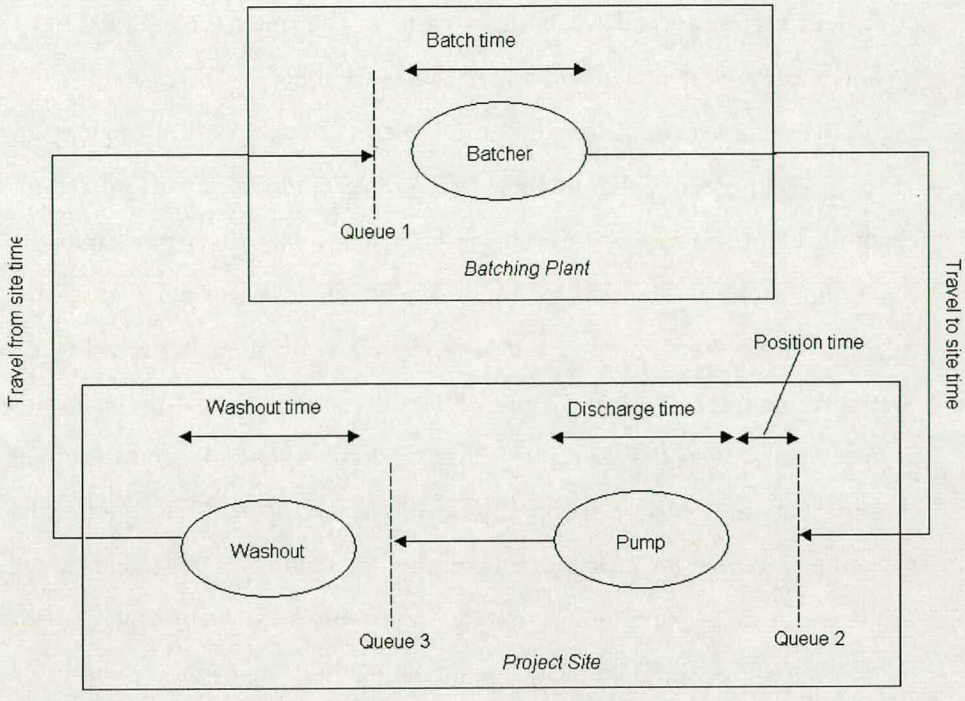


Figure 6.4: RMC supply process (with times of interest)

Initially the *Matlab* programming environment was used to develop this discrete-event simulation model, in a bid to make a powerful, flexible simulation model. The components of this model, *MatSim*, are as follows (the computer code written in the *Matlab* language is shown in appendix D):

- **Process state**: the state of the process is described in *MatSim* by whether the three servers (batcher, pump and washout) are idle or busy. For instance, if there is a truck mixer being batched with RMC at the batching plant, the batcher is in a busy state.
- **Simulation clock**: This variable gives the current value of simulated time, measured in seconds.
- **Input variables**: The user is required to provide some information to the *MatSim* model before simulation can be initialised. This includes the number of truck mixers to be used in the simulation, the number of each type of worker (e.g. concrete placer) to be used in the simulation, the stopping condition (e.g. volume of RMC to be placed), the probability distributions and their parameters.
- **Initialisation routine**: The simulation clock is set to zero; the process state is defined by truck mixers queuing at the batching plant; the batching plant being consequently

busy, and the pump and washout being idle. The times for the next events of each type to occur are calculated and added to the event list.

- **Event list**: This list comprises the times when the next event of each type will occur. There are six types of event possible in the *MatSim* model: a truck mixer arrival at the batching plant, pump or washout; and a truck mixer departure from the batching plant, pump or washout. The times for the next occurrence of each event are calculated from the random variables generated by the library routines (see later).
- **Statistical counters**. In *MatSim* the statistical counters record the amount of time (if any) that the servers (batcher, pump and washout) spend idle and the amount of time (if any) that truck mixers spend sitting idle in any of the three queues. Each time a truck mixer discharges a load into the pump, the volume of this load is noted.
- **Timing routine**: A subprogram that determines the next event from the event list and then advances the simulation clock to the time when that event is to occur.
- **Event routine**: A subprogram that updates the process state when a particular type of event occurs (there is one event routine for each event type).
- **Library routines**: A set of subroutines used to generate 1000 random observations (variates) from the probability distributions that represent each of the six times of interest. These 1000 random variates are averaged to provide one random variate (for each of the six times of interest) for use in the simulation model. The method of generating the random variates is probability distribution specific – thus an investigation into the types of probability distributions which adequately represent the six times of interest has been made and is detailed in section 6.5. To summarise the findings of this study, 7 distributions were found to be representative of the times of interest. They were: the uniform, normal, exponential, lognormal, weibull, gamma, and beta distributions. The methods for generating random variates from each of these distributions are detailed in Law and Kelton (2000).
- **Stopping condition**: This is the user-defined quantity of RMC that will result in the termination of the discrete-event simulation model.
- **Report generator**: A subprogram that computes estimates (from the statistical counters) of the desired measures of performance from the simulation run. These measures are as follows:
    - **Process duration (hours**
    - **Productivity (m3/hour)**

- **Direct cost (pounds)**
- **Truck mixer utilisation rate (%)**
- **On-site resources utilisation rate (%)**

- The details of how these measures are calculated in the model are discussed in appendix E.
- **Main program:** A subprogram that invokes the timing routines to determine the next event and then transfers control to the corresponding event routine to update the process state appropriately. The main program may also check for termination and invoke the report generator when the simulation is over.

These components work in the same fashion as those of the generic discrete-event simulation model as discussed in section 6.2.1.

6.4.1 Model verification

Verification is concerned with determining whether the conceptual simulation model (model assumptions) has been correctly translated into a computer program i.e. debugging the simulation computer program (Law and Kelton, 2000). Verification is an essential part of model development and one which should not be considered just in the latter stages of a project, but should be an elemental part of the development process. In this project the process of verification was integrated into the development process and was performed by organising the model into subprograms (which are easily debugged) and by debugging each sub program (e.g. timing routine) upon development. Additional verification was carried out by:

- Running the simulation under a variety of settings of the input variables, and checking to that the output was reasonable.
- Performing a "trace" on a simulation run. Where a trace involved comparing the state of the simulation, i.e. the contents of the event list, the state variables and the statistical counters, with some simple hand calculations of what the state of the simulation model should be, just after an event occurs.

The results from this verification process have shown that the model, *MatSim* was correctly modelling the RMC supply process.

6.5 Fitting probability distributions to the six times of interest

Simulation models are frequently driven by random variates that are generated from probability distributions that accurately represent the times of interest of a process. In this project there are six such times of interest and to produce an accurate discrete-event simulation model it is essential that the probability distributions (to be used) provide accurately representations of these times. Thus it is necessary to examine the collected data and assess which probability distributions are representative of the six times of interest.

To assist in this examination a bespoke computer application, *PDFit* was developed to allow probability density functions (PDFs) to be fitted to the collected data (Graham et al., 2005). It was thought the effort of developing a customised model was justified in this instance as the task is of the utmost importance in the development of a discrete-event simulation model and by developing a bespoke model the mathematical algorithms which fit the probability distributions to the data and assess this fit could be verified and the model output used with a high degree of confidence.

PDFit was developed using *Visual Basic for Applications* programming language, embedded into the *Microsoft Excel* spreadsheet package. It works by producing the PDFs of 27 select probability distributions plotted against the (six in this case) histogram(s) of the input data. The probability distributions considered in *PDFit* are: beta, binomial, chi-square, error function, erlang, exponential, extreme value, gamma, geometric, hyper geometric, inverse Gaussian, integer-uniform, logistic, log-logistic, lognormal, lognormal2, negative binomial, normal, pareto, pareto2, pearson type V, pearson type VI, poisson, rayleigh, Student's t, triangular, uniform and weibull. The mathematical forms of these distributions are provided by Law and Kelton (2000) and the Pallisade Corporation (2003).

The steps involved in *PDFit* to fit probability distributions to the input data are:

1. Create an input data histogram,
2. Estimate the parameters (e.g. mean) from the input data,
3. Plot PDFs (based on these parameters) against the input data histogram,
4. Visually assess the fit, and
5. Conduct goodness-of-fit tests.

The following sections discuss in detail how each of these steps are performed in *PDFit*.

6.5.1 Creating an input data histogram

To make a histogram, the range of values covered by the data is broken up into k disjoint adjacent intervals $[b_0, b_1]$, (b_1, b_2),.....(b_{k-1}, b_k). All the intervals should be the same width, say, $\Delta b = b_j - b_{j-1}$, which might necessitate removing extremely large or small X_i's to avoid producing an unwieldy-looking histogram plot. *For* $j = 1, 2, \ldots, k$ let h_j be the proportion of the X_i's that are in the jth interval (b_{j-1}, b_j). Finally, the function is defined as:

$$h(x) = \begin{cases} 0(zero) & \text{if } x < b_0 \\ h_j \text{ if } b_{j-1} \le x \le b_j & \text{for } j = 1, 2, \ldots, k \\ 0(zero) & \text{if } b_k \le x \end{cases} \quad (6.1)$$

and it is plotted as a function of x. The plot of h, which is piecewise-constant, is then compared with plots of densities of various distributions on the basis of shape alone (location and scale differences are ignored).

Histograms are applicable to any continuous distribution and provide a readily interpreted visual synopsis of the data. Furthermore, it is easy to visually judge the closeness-of-fit of density functions to the histogram. There are, however, certain difficulties with using histograms. Most irksome is the absence of a definitive guide for choosing the number of intervals, k (or equivalently, their width Δb).

The research literature contains numerous views on the optimum method of applying class interval rules. Montgomery and Runger (1994) proposed, in order to avoid uninformative histograms with too many or few class intervals, between five and twenty intervals should be used. They suggested that it is more accurate to calculate the number of intervals, k, by taking the square root of the number of observations, p.

This is known as the square-root rule, which is defined by:

$$k = \sqrt{p} \quad (6.2)$$

Their view is supported by Kottegoda and Rosso (1998), but the researchers extended the class intervals rule by setting a lower and upper bound of class interval quantity of 5 and 25, respectively. Neville and Kennedy (1964) suggested 10 and 25 as the lower and upper bounds, respectively.

Law and Kelton (2000) suggest using two alternative methods:

- Firstly, sturge's rule where the number of class intervals, k is defined by:

$$k = 1 + 3.322\log_{10}p \quad (6.3)$$

- Secondly, a variety of class intervals, k, should be used to produce a series of histograms. The histogram that has the smoothest shape should be utilized further.

It was decided upon consideration of the research literature that the square-root rule, sturge's rule and a selection of class interval values would be utilised in this study to decide the number of class intervals, k, upon which histograms would be formed. This resulted in k equalling 3, 6, 9, 12 and 15 from the selection; the values of k produced by the square-root and sturge's rules are presented next.

There were 2200 data points *(p)* for each of the times of interest, resulting in the number of class intervals defined by:

- Sturge's rule as $k = 1 + 3.322\log_{10}(2200) =\sim 12$
- Square-root rule as $k = \sqrt{2200} =\sim 47$

For each of these intervals (3, 6, 9, 12, 12, 15 and 47), a histogram was plotted and a visual comparison was undertaken to ascertain which histogram had the smoothest shape. The process of this visual assessment shall not be clarified here, rather it is illustrated in detail in 7.4.2. It was found that 12 class intervals produced the smoothest histogram for each of the six times of interest.

6.5.2 Parameter estimation

Probability distributions have at least two parameters which define them (Smith, 1998). There are a number of methods which can be utilised to estimate these parameters from the input data, such as maximum likelihood estimators (MLE), moment matching (MM) and least squares estimates. However, this discussion shall consider only MLEs further, as this method has been generally acknowledged to provide the most satisfactory estimates (Law and Kelton, 2000).

The likelihood function is $L(\theta) = f_\theta(X_1) \bullet f_\theta(X_2) \cdots f_\theta(X_n)$ (6.4)

(the notation of this equation can be found in appendix H). The MLE is defined to be the value of θ that maximizes $L(\theta)$ over all permissible values of θ..

6.5.3 Plot PDFs against input data histogram

The estimated parameters of a distribution allow the probability density function to be calculated. This PDF is plotted as a function of x (position time in figure 6.5).

There are many theoretical distribution functions which could fit a set of sample data, and because of this, the goodness-of-fit between the theoretical function and data should be assessed (Maio et al., 2000). One such method is a visual assessment. To allow a visual assessment of the closeness of fit to be undertaken, the PDF plot must be placed on top of the plot of the input data histogram. An example is shown in figure 6.5, where the PDFs of the lognormal and exponential distributions have been superimposed over the input histogram for the position time.

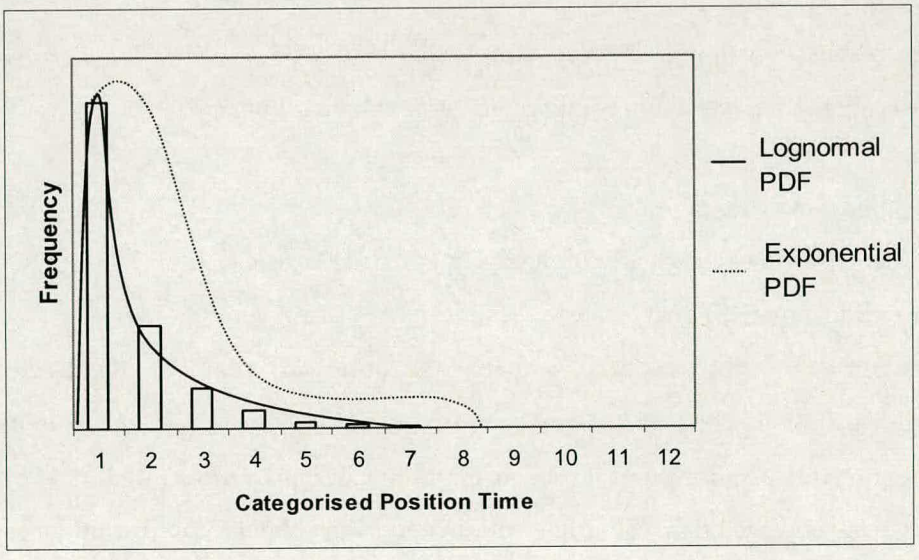


Figure 6.5: PDF vs. input data histogram (for position time) plot

6.5.4 Visual assessment of fit

To undertake a visual assessment of the goodness-of-fit, the shape of each PDF should be compared with the shape of the histogram. The *PDFit* program places the plots of PDF against histogram for each distribution together, to aid the comparison process. The PDF that has the closest matching shape, should be selected and used to represent the input data in any further analysis. For example in figure 6.5 the PDFs of the lognormal and exponential distributions

have been plotted against the input histogram of position time. A visual assessment of the fit of these PDFs would herald that the lognormal distribution provides a better fit to the input data than the exponential distribution. This is because the lognormal PDF has a similar shape to the input data, where the exponential does not.

Abourizk et al. (1991) expressed the opinion that visual assessment of the fit is often the superior method of assessing the closeness-of-fit. However, visual assessment will always be subjective and prone to potential human error and therefore mathematical support for the assessment can improve confidence in the selection. *PDFit* includes two formal statistical tests to provide such mathematical support. They are the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests. A common alternative statistical assessment of fit is the Chi-Square test. Although the Chi-square test was used by Clemmens and Willenbrock (1978), it has not been included in *PDFit* because the KS and AD tests have several advantages:

- There is no requirement for grouping of data;
- The test results are valid for the exact sample size; and
- The tests are more powerful in comparing alternative distributions.

6.5.5 Goodness-of-fit tests

Two formal goodness-of-fit tests are conducted in PDFit. They are:

Kolmogorov-Smirnov (KS) Test

The Kolmogorov-Smirnov (KS) test compares the observed data with the hypothesized theoretical distribution. This test does not require the grouping of data and eliminates the problem of interval specification (a problem with the Chi-Square test) (Maio et al., 2000). It also checks if the observed data could have originated from a theoretical distribution, with the estimated parameters (Maio et al., 2000).

$$F_n(x) = \frac{number\ of\ X_i's \leq x}{N} \quad for\ all\ real\ numbers\ x \tag{6.5}$$

Thus, $F_n(x)$ is a (right-continuous) step function such that

$$F_n(X_{(i)}) = \frac{i}{n} \tag{6.6}$$

for $i = 1, 2, \ldots, n$. If $F(x)$ is the fitted distribution function, a natural assessment of goodness-of-fit is a measure of the closeness between $F_n(x)$ and $F(x)$. The KS test statistic D_n is simply the

largest (vertical) distance between $F_n(x)$ and $F(x)$ for all values of x and is defined formally by (notation is in appendix H):

$$D_n = \sup_x \{|F_n(x) - F(x)|\} \tag{6.7}$$

Abourizk et al. (1991) expressed the opinion that the KS statistic can be used for assessing the quality of fit from various methods of statistical analysis e.g. fitting theoretical distributions, empirical distributions, etc. Abourizk et al. (1994) used the KS test on construction process data sets.

Anderson-Darling (AD) Test

The Anderson-Darling (AD) test differs from the KS test on the weight given to the tails of the distributions (Maio et al., 2000). In fact, the AD test is designed to detect discrepancies in the tails of probability distributions (Law and Kelton, 2000).

The AD test statistic A_n^2 is defined by (notation in appendix H):

$$A_n^2 = n\int_{-\infty}^{\infty}\left[F_n(x) - F(x)\right]^2 \Psi(x) f(x)dx \tag{6.8}$$

Thus, A_n^2 is just the weighted average of the squared differences $[F_n(x) - F(x)]^2$ and the weights are the largest for $F(x)$ close to 1 (right tail) and $F(x)$ close to 0 (left tail).

It was suggested that the AD test has higher power than the KS test against many alternative distributions (Stephens, 1974). The AD test has not, to date, been used extensively in construction-based research (Maio et al., 2000).

6.5.6 Assessment of the fit of PDFs to the times of interest

To recap there are 186 (operations) data points and each of these contains a varying number of observations of the six times of interest, as there is a time of interest recorded for each truck mixer delivery within an operation. 27 probability distributions were fitted to each of the times of interest in the 186 operations. These fits were assessed using a visual method, the KS test and the AD test and the probability distribution which provided the best fit (as identified by each of these methods) was recorded for each of the times of interest in all 186 data points. It was found that only 7 of the 27 distributions provided a best fit for one of the 186 data points, thus the remaining 20 distributions have been excluded from the plot.

The results of the assessment of the travel to site time of interest are shown in figure 6.6. Clearly in a significant number of operations the lognormal, gamma and beta distributions provided the best fit. Depending on which statistic is given the most prominence, then the lognormal distribution could be chosen to provide a representation of this time of interest (from visual assessment or AD test); or the gamma distribution may provide a good overall representation of the travel to site time (from KS test).

The results of the assessment for the remaining times of interest are displayed in graphs F.1 to F.5 in appendix F. The overwhelming trend from these results is that no one probability distribution is capable of providing a representation of any of the times of interest. Although the option of which probability distribution should be used to represent a particular RMC operation has been reduced to a choice of one out of seven, this is still a complex task for a construction practitioner with little or no statistical knowledge of the process. Coupled with this problem is the need to select appropriate values for the parameters of the chosen distribution. The parameters of a distribution dictate the shape of the PDF of the distribution and hence govern the values of the estimates taken from the distribution in generating random variates (see above). Thus, without such parameter values the simulation model will not be able to recreate the process accurately, making simulation futile.

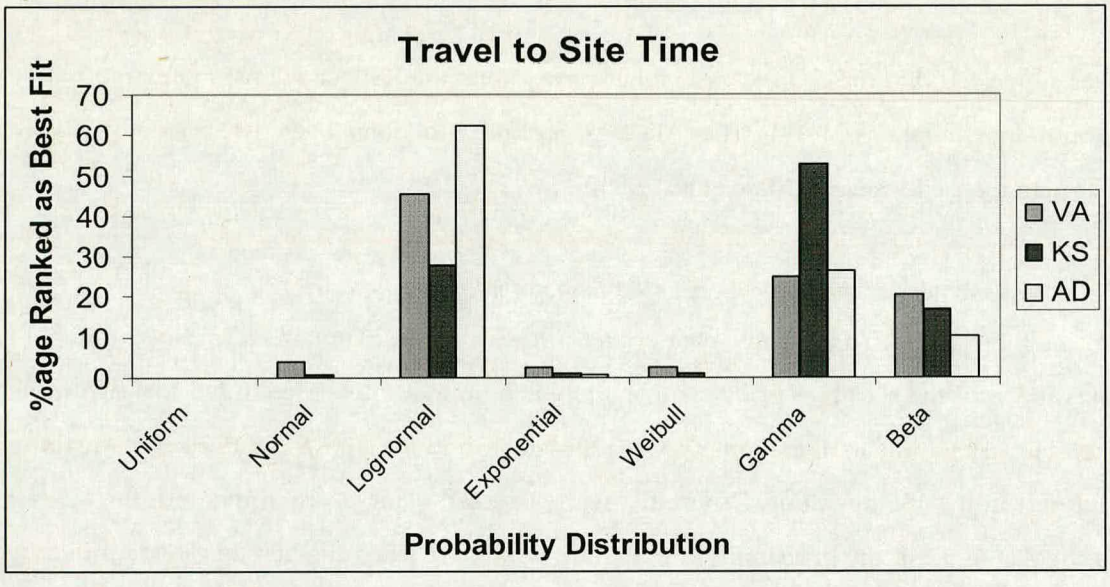


Figure 6.6: Assessment of fit of PDFs to the travel to site time data

6.6 Model validation

Validation is the process of determining whether a model is an accurate representation of the process, for the particular objectives of a modelling study (Law and Kelton, 2000). Naylor (1971) proposed a three-stage validation procedure that has been adopted by many modellers of economic and scientific systems and processes. These stages are:

- Formulation of the hypotheses and postulates for the process under study using the available knowledge such as observations, relevant theory or even intuition. Law and Kelton (2000) proposed the expansion of this stage by considering the face validity of the model. Face validity is whether or not a model appears to be a reasonable solution to those who are knowledgeable about the process.
- Validation of the assumptions upon which the model is based by subjecting them to empirical testing.
- To compare the model output with real-world observations, the validity of a model is demonstrated, if it has the ability to predict the behaviour of the real processes variables, irrespective of whether or not the assumptions the model is based upon, are reasonable. This is the most definitive test of a simulation model's validity (Law and Kelton, 2000) and has thus been made the focus of this study.

MatSim is intended to represent the existing process, not replace it. The model is also intended to allow estimations of the process output (performance measures) to be produced. Therefore, the model output and data from the existing system can be compared to ascertain if the model is replicatively valid. Naylor (1971) suggested using classical statistical tests such as analysis of variance (ANOVA) or the Chi-square test, to determine the goodness-of-fit between the model and the existing (observed) process output. The tests are not applicable in this case because discrete-event simulation output data (for a single time series) is usually highly auto-correlated (Law and Kelton, 2000). Thus, a different approach must be adopted and in this study, and in this study the correlated inspection approach was used to indicate the replicative validity of the *MatSim* model. Correlated inspection at its most extreme requires that the model be driven by exactly the same observations as the real process. This pre-supposes that such data from a historical process is available, but it is difficult to argue whether or not this indicated the predictive powers of a model (Smith et al., 1995). Therefore the model will be

driven with random variates generated from probability distributions (and their parameters) chosen by the author (in this validation study).

The details of 50 RMC operations were retained for use in the validation process and the author selected a probability distribution (and parameters) to represent each of the six times of interest given the specific details of each of the 50 operations. The operations were then recreated using *MatSim* and the estimates of the process performance measures were recorded for each operation. These estimates were then compared directly with the observed process performance measures of the 50 operations, in a correlated inspection approach to validation. A summary of this correlated inspection is shown in figure 6.7.

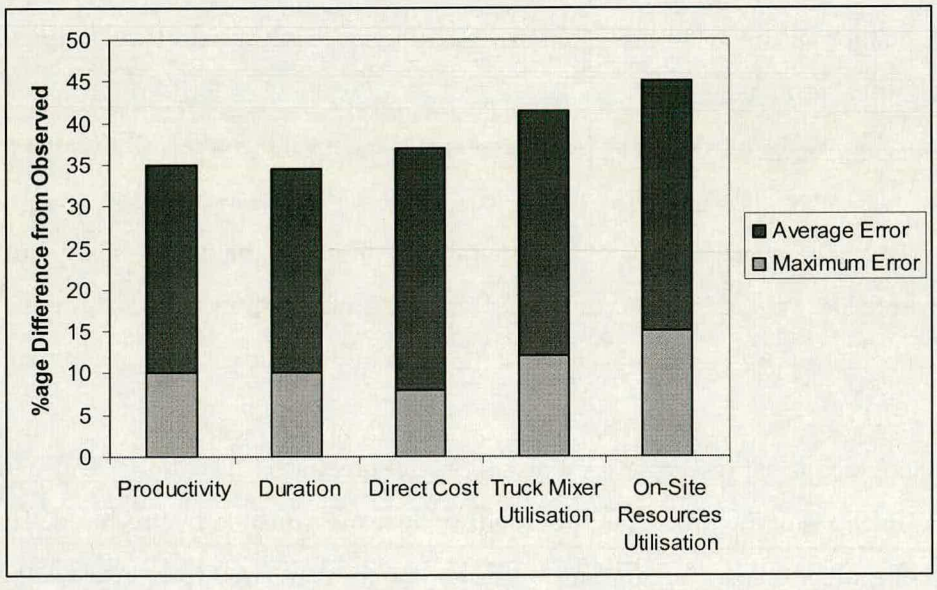


Figure 6.7: Correlated inspection results from the validation of *MatSim*

The correlated inspection of *MatSim* revealed that the model is not capable of producing accurate estimates of the RMC supply process performance measures. It is felt that this is because this inputs to the simulation model are of a poor quality. These inputs being the probability distributions and their parameters which are used to represent the process and to generate estimates of the duration of the individual tasks involved in the process. The decision on which probability distributions and especially on what the parameters of the chosen distributions is not an easy one. Certainly it was difficult for the author (who has been studying statistics and process modelling for 3 years) to make this decision for each of the operations in the validation study. It is not outrageous to assume that a construction

practitioner who has not had the benefit of such study would find this problem even more complex. This problem shall be coined *'the complexity problem'* in the remainder of this thesis and solution to this problem is required to permit construction practitioners to access the benefits of using a simulation model, such as easily performed experiments to aid decision-making. Such a solution shall be explored in the next chapter.

6.7 Chapter conclusions

The following conclusions may be drawn from the discussion contained in chapter 6:

- Poor performance by the UK construction industry may be partially attributed to the poor performance of cyclical, stochastic construction processes. This poor performance is caused by an inadequate allocation of resources to construction operations, due to intuitive estimate of the required resources. This chapter focused on developing and validating a discrete-event simulation model to remove the need for a construction practitioner to make such intuitive estimates.
- A discrete-event simulation model was developed using a general programming language *Matlab*, to enable the author to learn a great deal about the inner workings of the model. This model was verified and validated.
- The above model was named *MatSim*. This model consists of 3 FIFO queues, representing the possible queues that could form in the RMC supply process at the batching plant, concrete pump and washout area. In the model a specified number of truck mixers travel around the RMC supply cycle, queuing if necessary at the 3 mentioned locations. In the model the RMC supply process can be viewed as consisting of 6 activities: batching a load; a truck mixer travelling to the project site with the load; positioning a truck mixer at the concrete pump; discharging the load into the hopper of the concrete pump; washing out the residue from an empty truck mixer; and a truck mixer travelling back to the batching plant to receive another load. The durations of these activities were termed batch time; travel to site time; position time; discharge time; washout time and travel from site time – collectively known as the *'times of interest'* in the model. These times are of interest because they are non-deterministic and represent most of the stochastic nature of the RMC supply process. By estimating the value of all 6 *times of interest* it is possible to build up a typical and accurate picture of the cycle of a truck mixer as it travels from the batch plant to a project site and back again to the batch plant; and by building up multiple cycles it is

possible to estimate the process performance measures. The estimates of the *times of interest* are made from probability distributions that accurately represent these times.

- The probability distributions that accurately represent the *times of interest* were then investigated. This was done by fitting theoretical probability density functions to the histograms produced by the observed data (for each time of interest) using a bespoke computer package, *PDFit*. Such a package was developed in this study to allow the mathematical functions used in the model to be verified.
- Three methods of assessing the fit of a probability density function to the above histograms were used: visual assessment, the Kolmogorov-Smirnov test and the Anderson-Darling test. The findings of all three methods agreed that no one probability distribution adequately represents all of the *times of interest*, and the distribution (and its parameters) will differ for each *time of interest* given differing operating conditions.
- This resulted in a problem – a construction practitioner is expected to possess enough statistical knowledge of the RMC supply process to choose the probability distribution (and its parameters) that represent the *times of interest*. This problem was termed *'the complexity problem'*. *The complexity problem* makes MatSim a difficult to use, impractical model and hence it is an issue that must be resolved. A solution to *the complexity problem* is investigated and proposed in chapters 7 and 8.
- A measure of the significance of *the complexity problem* was presented in the validation study. The authors choose the probability distributions (and their parameters) that represent the *times of interest* of the RMC supply process for 50 past RMC operations. The estimates (of process productivity) produced by MatSim (of these 50 operations) were not very promising – they were often distinctly different from the observed process productivity.

CHAPTER 7

A SOLUTION TO *THE COMPLEXITY PROBLEM*

It has been established that the correct probability distribution, as input to a simulation model, is paramount. Unfortunately, the typical construction practitioner rarely possesses detailed knowledge of complex distributions and this *complexity problem* presents a barrier to both the quality of such a model and its take-up within the industry. This chapter introduces *Case-Based Reasoning* and its use in solving this problem.

7.1 Introduction

The simulation model presented and validated in chapter 6 was declared to be too complex to be practical. The specific problem is that a user of the simulation model is required to be able to choose the probability distributions and parameters that represent the *times of interest* contained within the RMC supply process; this is not an easy task unless the user has expert knowledge of both the process and of probability distributions in general. This problem with the *MatSim* model was dubbed *the complexity problem* and a method shall be sought here to reduce – or remove – this problem, and make the model practicable.

To overcome *the complexity problem* an artificial intelligence (AI) method, case-based reasoning (CBR) has been utilised to suggest the probability distributions that represent a process given a specific set of operating conditions, thus removing the need for a user to perform this task. These suggested distributions are then supplied to a simulation model, which can then provide estimates of the process performance measures using the same method described in chapter 6. With this method a hybrid model is born; CBR provides the input to a simulation model, which subsequently provides the output to the user. This model is known as *CBRSim*.

CBR has been considered over other AI methods, such as Neural Networks, because of a CBR model's ability to learn well from, and hence recreate, past cases when there is a well-defined problem (Arditi and Tokdemir, 1999; Forbes et al., 2004), and this is the type of problem that is being dealt with in this research project.

The development of CBR models is mainly a heuristic process and before *CBRSim* can be developed the main parts of a CBR model – the indexing, retrieval, adaptation and learning mechanisms – need to be optimised as far as possible. This chapter discusses the optimisation of the key parts of a CBR model. The *CBRSim* model is comprised of a CBR module with these optimised parts and a simulation model and is presented in chapter 8.

In this chapter the following is discussed:

- **Section Two.** What CBR is and why it has been used in this project is examined. This includes a literature review of the extent of use of CBR in general prediction and estimation problems.

- **Section Three.** The details of the case base developed in the CBR models are presented.
- **Section Four.** There is no fixed theory on how a CBR model should be developed. Rather, this process is usually performed in an ad-hoc manner. Here, different methods are considered for performing the indexing of cases.
- **Section Five.** Methods for performing case retrieval (including a novel method of performing this task) are discussed. Also presented is how variable weightings were calculated (for use in similarity measurement), and an examination is made of the sensitivity of the case retrieval mechanism to different values of variable weighting.
- **Section Six.** A number of models were developed using various combinations of the indexing and retrieval methods described in 7.4 and 7.5, these models are discussed in this section and are then validated. The combination of indexing and retrieval mechanism that produced the most accurate model was chosen for use in the hybrid model, *CBRSim*. This model is then presented in chapter 8.
- **Section Seven.** Chapter summary and conclusions.

7.2 What is case-based reasoning (CBR) and why is it used in this project?

Case-based reasoning (CBR) is an artificial intelligence technique whose origins lie in the work of Shanks and Abelson (1977). It is the process of reasoning and learning by storing cases in a database (case base) – which are the relevant details (input variables and output probability distributions) of one construction operation in this model – and retrieving and adapting them to aid new problem solving or interpretation in similar situations (Kolodner, 1993; Leake, 1996).

A CBR model consists of a number of parts (shown in figure 7.1):

- **Indexing mechanism.** Watson (1997) defined a case as 'a contextualised piece of knowledge representing an experience'. It contains the past lesson that is the content of the case and the context in which the lesson can be used. A case can be an account of an event, a story, or some record - in this instance, the process performance measures observed in a particular operation of the RMC supply process - typically comprising:
    - The problem that describes the state of the world when the case occurred, i.e. the set of input variables observed during a concreting operation.

- The solution that states the derived solution to that problem, i.e. the process performance measures observed during a concreting operation.

An indexing mechanism fulfils the function of ensuring that each raw case or new problem to be solved (target case) is presented to the CBR model in a standardised form.

- **Case base.** The case base is the storage area for prior cases.
  A robust case base, containing a representative and well-distributed set of cases is the foundation of a good CBR model (Kriegsman and Barletta, 1993).
- **Retrieval mechanism**. A retrieval mechanism compares the target case (the problem to be solved) with cases stored in the case base, using some form of similarity (distance) measurement. This measurement may add weightings to the case variables according to their relative importance, or not. After comparing the cases this mechanism suggests the output of the most similar case (least distance between this stored case and the target case) as a solution to the target case (problem). This may not be the final answer produced by the CBR model – some tinkering or adaptation – may be required (see next bullet point).

A CBR system derives its power from its ability to retrieve relevant cases quickly and accurately from its case base (Arditi and Tokdemir, 1999), and is thus the retrieval mechanism is a very important part of a CBR model. The work in this thesis is not reliant on a particular retrieval mechanism, rather a number of mechanisms have been surveyed and compared to reveal which works best in this particular application.

- **Adaptation mechanism**. An adaptation mechanism adapts the solution suggested by the retrieval mechanism to the needs of the target case using a set of mathematical rules. While adaptation is useful in many situations, it is by no means essential (Watson, 1997), and as there are no easily defined mathematical rules to govern the relationship between the process performance measures and the operating conditions in the RMC supply process, adaptation has been ignored in this project.
- **Learning mechanism.** CBR is an artificial intelligence technique because it possesses the capacity to learn of its own volition. This practice of learning is performed by a learning mechanism. This mechanism is usually a simple one, which takes the input details (values of the variables) of the target case and the (adapted) output from the

most similar case (see retrieval mechanism above) and makes a new case from these inputs and outputs. This forms the final solution, which is not only presented to the user of the model, but is stored in the case base as a new case, thus providing the learning capacity of CBR.

When the above parts of a CBR model have been developed, the process involved of obtaining a solution to a problem is cyclical (Aamodt and Plaza, 1994) this can be seen in figure 7.1. This process involves:

- Retrieval of the most similar case from the case base.
- Revision (adaptation) of the solution if necessary.
- Re-use of the (adapted) case to solve the problem.
- Retention of the final case as a new case in the case base.

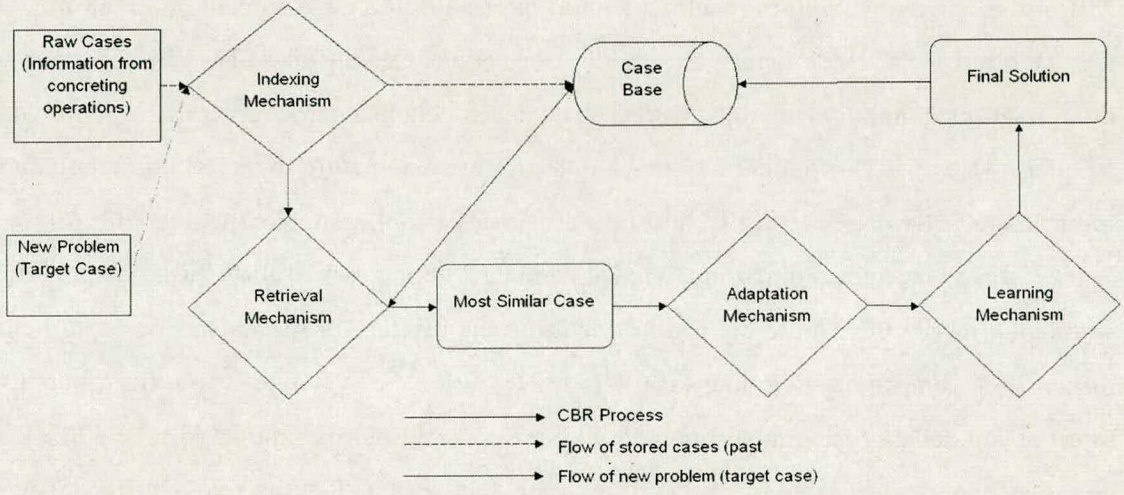


Figure 7.1: The parts of a CBR model and the interaction between them

Has CBR been utilised in the construction domain, and in particular to solve estimation and prediction problems (like the one presented here)? The answer is yes. The construction domain is a growing area for CBR applications, particularly in the area of design (Tah et al., 1999). Such applications include: CASETOOL, a system for bridge design (Kumar and Krishnamoorthy, 1995) and BRIDGER, a system to aid the conceptual design of cable-stayed bridges (Reich and Fenves, 1995). Faltings et al. (1991) and Shih (1991) applied CBR to architectural design. Flemming (1994) and Rivard et al. (1988) addressed building design issues. Roddis and Bocox (1997) created a system to resolve fabrication errors in steel highway

bridges. In the area of planning, Tah et al. (1999) created CBRidge Planner to aid the planning of highway bridge projects. CasePlan was designed for scheduling and planning boiler erection by Dzeng and Tommelein (1997). All of the aforementioned CBR systems design, diagnose or plan - but they do not predict. In the past, CBR has been used successfully in solving prediction and estimation (Morcous et al., 2002). In construction, Arditi and Tokdemir (1999) have investigated the application of CBR to predicting construction litigation outcome.

Clearly there have been a number of CBR models relating to construction and civil engineering, some of which have successfully solved prediction and estimation problems. Hence, it is likely that CBR could provide a solution to the prediction-type problem that is being addressed in this project. This prospect shall be explored further in this chapter.

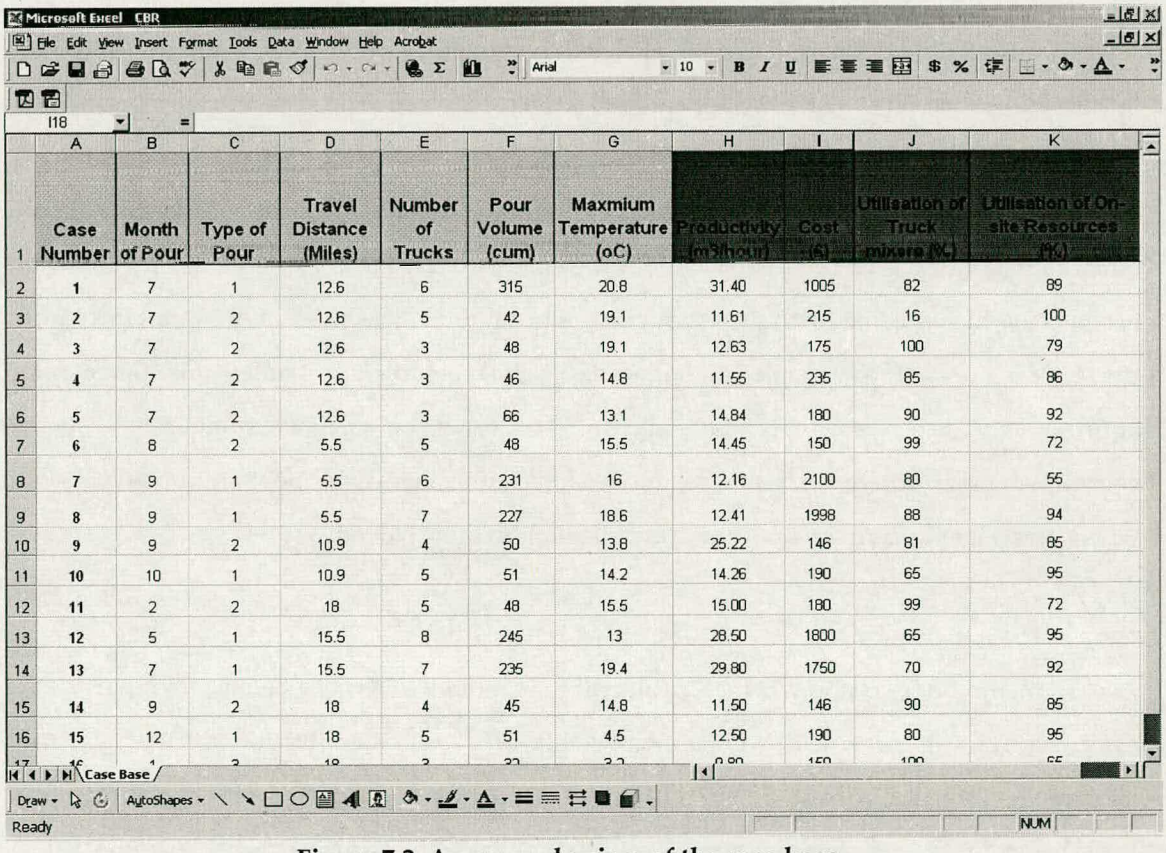
A decision on what computing method should be used in the development of a CBR model was required to be made before any model development could take place. In chapter 6 a discrete-event simulation model, *MatSim* was developed using the general programming language *Matlab*. A CBR model requires a database of some nature to store information and unfortunately the database available (*Microsoft Access*) is not easily compatible with *Matlab*. Thus, if there was to be any requirement for communication between the CBR model and the simulation model this would be made difficult by the inability of *Matlab* and *Access* to hold discourse. A compromise solution was found in *Microsoft Excel*, which possesses the capability to act as a database (using tables) and could be used to develop a simulation model like the one developed in *Matlab* by using *Visual Basic for Application (VBA)* macros. All CBR models that were developed in this project were done so using *Microsoft Excel*. To permit easy communication with the CBR models, the simulation model *MatSim* was re-developed in *Microsoft Excel*. This new model is known simply as *ExcelSim*, and the source code of this model may be found in appendix D. To ensure that this new model is valid, the validation process (detailed in 6.6) was repeated and the results are shown in appendix G. To summarise, *ExcelSim* was found to be as valid as *MatSim*.

7.3 Case base design

In this project the information from 232 concreting operations is available. However, all of this information shall not be stored in the case base. Instead the data shall be split using a ratio of

80:20, with 80% of the concreting operations being used in model development, resulting in a case base of 186 cases. The other 20% or 46 cases shall be reserved for testing (or validating) the models discussed in section 7.6.

In chapter 5 a set of variables was defined for use in any model development in this PhD project. These variables are statistically significant and likely to be known by a construction practitioner at the time of placing an order for RMC (order time). Here all of the models discussed in section 7.6 have been developed using these variables, resulting in a case having six input variables and four output variables (productivity, cost, utilisation of truck mixers and utilisation of on-site resources). An example of a case base consisting of 15 cases of this description is shown in figure 7.2.



	A	B	C	D	E	F	G	H	I	J	K
1	Case Number	Month of Pour	Type of Pour	Travel Distance (Miles)	Number of Trucks	Pour Volume (cum)	Maxmium Temperature (oC)	Productivity (m3/hour)	Cost ($)	Utilisation of Truck mixers (%)	Utilisation of On-site Resources (%)
2	1	7	1	12.6	6	315	20.8	31.40	1005	82	89
3	2	7	2	12.6	5	42	19.1	11.61	215	16	100
4	3	7	2	12.6	3	48	19.1	12.63	175	100	79
5	4	7	2	12.6	3	46	14.8	11.55	235	85	86
6	5	7	2	12.6	3	66	13.1	14.84	180	90	92
7	6	8	2	5.5	5	48	15.5	14.45	150	99	72
8	7	9	1	5.5	6	231	16	12.16	2100	80	55
9	8	9	1	5.5	7	227	18.6	12.41	1998	88	94
10	9	9	2	10.9	4	50	13.8	25.22	146	81	85
11	10	10	1	10.9	5	51	14.2	14.26	190	65	95
12	11	2	2	18	5	48	15.5	15.00	180	99	72
13	12	5	1	15.5	8	245	13	28.50	1800	65	95
14	13	7	1	15.5	7	235	19.4	29.80	1750	70	92
15	14	9	2	18	4	45	14.8	11.50	146	90	85
16	15	12	1	18	5	51	4.5	12.50	190	80	95

Figure 7.2: An example view of the case base

7.4 Indexing mechanism

To restate, an indexing mechanism fulfils the function of ensuring that each raw case or new problem to be solved (target case) is presented to the CBR model in a standardised form. This

section outlines the two methods considered for indexing cases - crisp and fuzzy, and provides details of how the sizes of indices are defined in each case.

In CBR research to date, most models make the assumption that the data being dealt with is of a qualitative type (Jeng and Liang, 1995) and the indexing of such data types usually involves ensuring that the same qualitative descriptions are used to describe the same things over the entire case base. In the data being used in this project two variables are qualitative - they are the month of pour and type of pour. These values are simply recorded in the case base as they were observed e.g. a concreting operation of a base-type pour taking place in September, will be recorded as having: base as the type of pour, and September as the month of pour, in the case base. These methods are standard across all of the models produced in this study.

There has been little discussion in general as to how quantitative-type data should be indexed in CBR models, and this certainly appears to be the case in construction and civil engineering related research. This is clearly a problem which needs to be solved in CBR modelling of cyclical, stochastic construction processes, where the majority of the observed data is quantitative (four out of six variables in this project). To overcome this problem, a method has been used which is common in the production of histograms. This method examines the entire range of values of a variable contained in a data set then separates this range into appropriately sized categories (or indices). Two types of category (index) have been considered: crisp, where a variable's value is either a member of an index or not; and fuzzy, where a variable's value can hold some membership to multiple indices.

7.4.1 Crisp method

A crisp method of defining whether a value of a quantitative variable belongs to a particular category (index) is governed by crisp membership functions. These functions are simple and usually are of the form that if a particular value falls within the range of a particular category then this value is a member of that index; if it falls outside the particular range then this value is not a member of that index. To illustrate this concept, an example of such crisp membership functions is shown in figure 7.3.

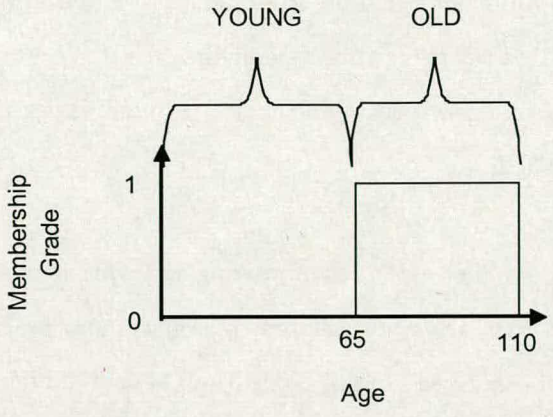


Figure 7.3: Example crisp membership functions

If a person is required to be classified or indexed according to age, then crisp membership functions such as those shown in figure 7.3 may be used. To state these functions in mathematical form, they are:

$$Membership\ (person) = \begin{cases} old & if\ person\ is\ aged\ 65\ to\ 110 \\ young & if\ \ person\ is\ aged < 65 \end{cases} \tag{7.1}$$

According to the crisp membership functions contained in equation 7.1 and displayed in figure 7.3, a person less than 65 years of age can be defined as *young*, and anyone over 65 is declared as *old*.

In the CBR models developed here using a crisp indexing method, the mechanism simply places the value of a variable contained in a raw case (or target case) into the appropriate index and records the number of this index. This number now becomes the value recorded in the case base for that particular variable for that particular case. For example, if the smallest pour observed was of volume $10m^3$ and the largest was $300m^3$, the range of total pour volumes is 10 – 300 m^3. Furthermore if (using some method) an appropriate sized index for total pour volume was defined as 50 m^3, the indices and their corresponding index numbers are as follows: 10 – 59 m^3, index 1; 60 – 109 m^3, index 2; 110 – 159 m^3, index 3; 160 – 209 m^3, index 4; 210 – 259 m^3, index 5; 260 – 309 m^3, index 6. If the total pour volume of a case was 215 m^3, then the index mechanism would place this value in index 5, and 5 would now become the value for total pour volume in this particular case.

A crisp indexing mechanism will ensure that all of the data presented in the case base is of a standard form and has the added advantage of reducing the area that the retrieval mechanism must work on, e.g. searching integers between 1 and 6 is much easier than searching real numbers over a range of 10 to 300.

The main concern with the above method is how to define appropriate crisp index sizes. This difficulty is common and has been acknowledged by researchers and practitioners who have utilised histograms in the past (Law and Kelton, 2000). Their general advice is to try a number of different methods to define appropriate index sizes, and that has been done here. The details of these different methods are discussed in 7.4.2.

7.4.2 Defining the crisp index sizes

In this project three methods for defining crisp index sizes have been considered. These are (from Law and Kelton, 2000):

- **Sturge's rule**

Sturge's rule is commonly used to define the number (or alternatively the width) of indices in the formation of histograms. The governing equation of Sturge's rule is:

$$k = 1 + 3.322 \log_{10} p \tag{7.2}$$

where p is the number of data points (186 in this instance) and k is the number of indices that should be made.

When p = 186, k = 1 + 3.322 x $\log_{10}$ (186) = 8.6 ~= 9 indices. For example, this results in the variable total pour volume (which has a range from 10 m^3 to 282 m^3) being indexed as follows: 10 – 45.7 m^3, index 1; 45.8 – 79.3 m^3, index 2; 79.4 – 113 m^3, index 3; 113.1 – 146.7 m^3, index 4; 146.8 – 180.3 m^3, index 5; 180.4 – 214 m^3, index 6; 214.1 – 247.7 m^3, index 7; 247.8 – 281.3 m^3, index 8; 281.4 + m^3, index 9. The indexing mechanism produced by Sturge's rule is summarised in figure 7.4.

- **Square-root rule**

The square-root rule, like Sturge's is commonly used in histogram formation. The governing equation of the square-root rule is:

$$k = \sqrt{p} \tag{7.3}$$

When p = 186, k = √186 = 13.6 ~= 14 indices. For example, this results in the variable total pour volume (ranged from 10 m^3 to 282 m^3) being indexed as follows: 10 – 32.2 m^3, index 1; 32.3 – 52.4 m^3, index 2; 52.5 – 72.6 m^3, index 3; 72.7 – 92.8 m^3, index 4; 92.9 – 113 m^3, index 5; 113.1 – 133.2 m^3, index 6; 133.3 – 153.4 m^3, index 7; 153.5 – 173.6 m^3, index 8; 173.7 – 193.8 m^3, index 9; 193.9 – 214 m^3, index 10; 214.1 – 234.2 m^3, index 11; 234.3 – 254.4 m^3, index 12; 254.5 – 274.6 m^3, index 13; 274.7+ m^3, index 14. The indexing mechanism produced by the square-root rule is summarised in figure 7.5.

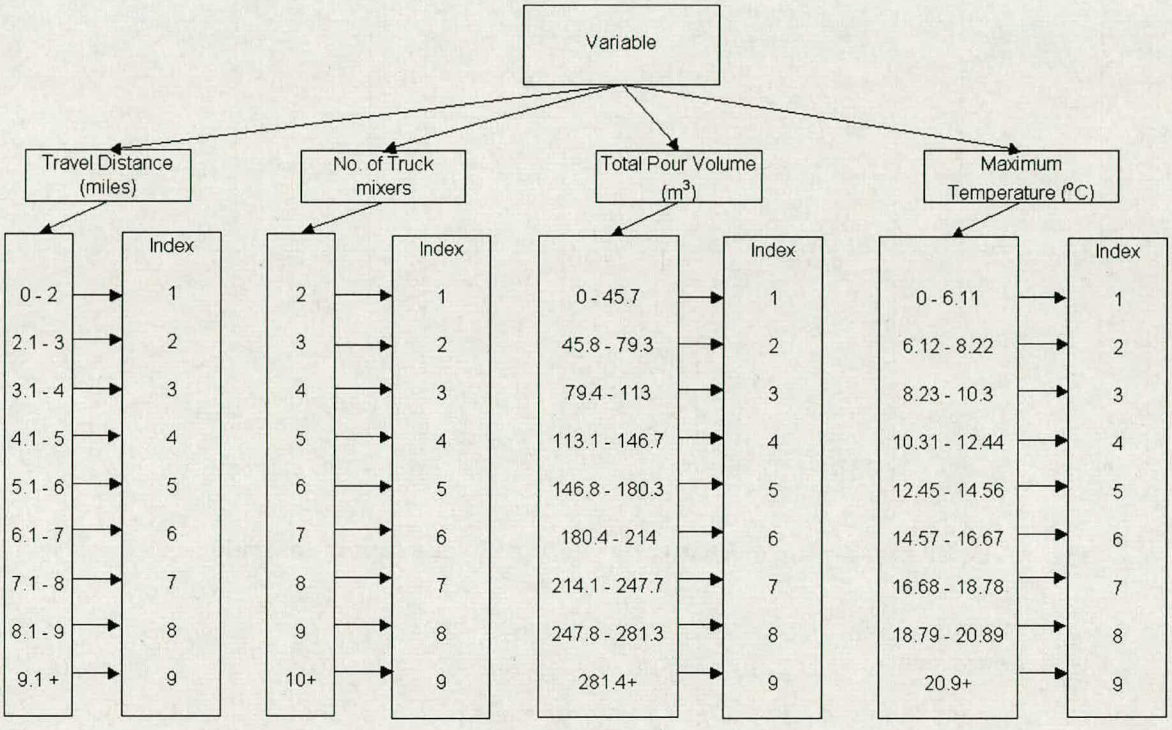


Figure 7.4: Crisp indexing mechanism as defined by Sturge's rule

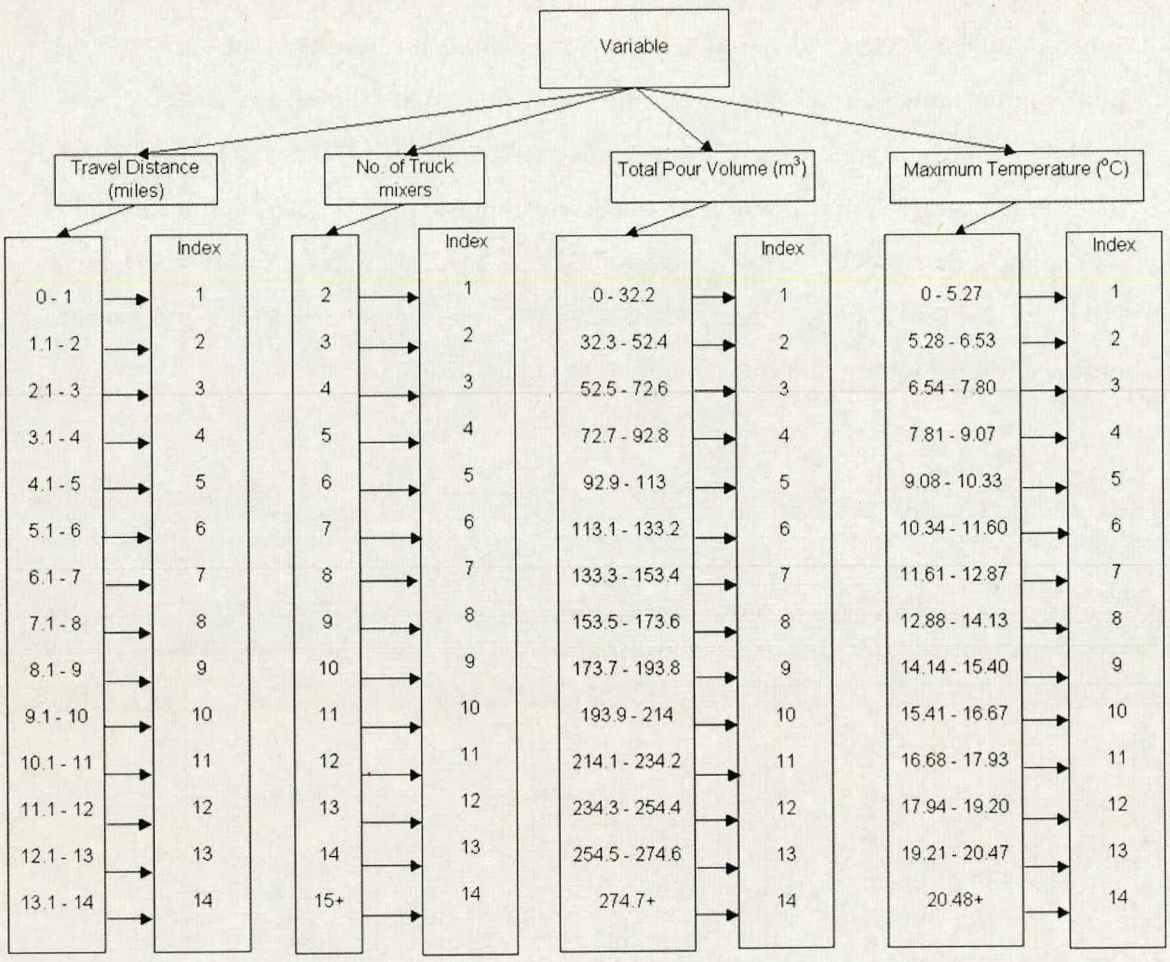


Figure 7.5: Crisp indexing mechanism as defined by the square-root rule

- **Visual assessment**

Visual assessment involves: choosing a range of the number of indices in which to separate (or index) the quantitative variables e.g. the number of indices could vary from 3 and 15 in increments of 3; plotting the frequency graph produced from the data set for each of these numbers of indices; and finally selecting the number of indices that produces the smoothest frequency graph. Visual assessment has been considered here as suggestion has been made in the past that it is the superior method of determining the sizes of indices in histogram formation (Abourizk et al., 1991).

Each of the quantitative variables was indexed (separated) into 3 indices and a frequency graph (e.g. figure 7.6a) was produced. This number of indices was increased in increments of 3 until each variable was separated into 15 indices, and a frequency graph was

produced for each of these numbers of indices. This resulted in five frequency graphs for each variable (such as those shown for travel distance in figures 7.6a to 7.6e). An example of this step is shown below:

The quantitative variable travel distance was separated into 3, 6, 9, 12 and 15 indices. The resultant ranges of these indices in each of these cases are shown in table 7.1.

	Number of Indices				
Index Number	3	6	9	12	15
1	0 - 5	0 - 2	0 - 1	0 - 1	0 - 0.75
2	5.1 - 10	2.1 - 4	1.1 - 2	1.1 - 2	0.76 - 1.5
3	10.1+	4.1 - 6	2.1 - 3	2.1 - 3	1.51 - 2.25
4	~	6.1 - 8	3.1 - 4	3.1 - 4	2.26 - 3.00
5	~	8.1 - 10	4.1 - 5	4.1 - 5	3.01 - 3.75
6	~	10.1+	5.1 - 6	5.1 - 6	3.76 - 4.5
7	~	~	6.1 - 7	6.1 - 7	4.51 - 5.25
8	~	~	7.1 - 8	7.1 - 8	5.26 - 6.00
9	~	~	8.1+	8.1 - 9	6.01 - 6.75
10	~	~	~	9.1 - 10	6.76 - 7.5
11	~	~	~	10.1 - 11	7.51 - 8.25
12	~	~	~	11.1+	8.26 - 9.00
13	~	~	~	~	9.01 - 9.75
14	~	~	~	~	9.76 - 10.5
15	~	~	~	~	10.51+

Table 7.1: Indices defined by visual assessment for travel distance (all in miles)

The travel distance data was separated into the indices defined in table 7.1, and a number of frequency graphs were produced. For the case when the quantitative data was indexed into 3 indices, the frequency graph produced is shown in figure 7.6a. The frequency graphs produced by the remainder are shown in figures 7.6 b to e. A visual assessment to identify the number of indices that produces the smoothest frequency graph is made simple by drawing a line of smoothness on each graph. Clearly the smoothest frequency graph is produced when travel distance was separated into 3 indices.

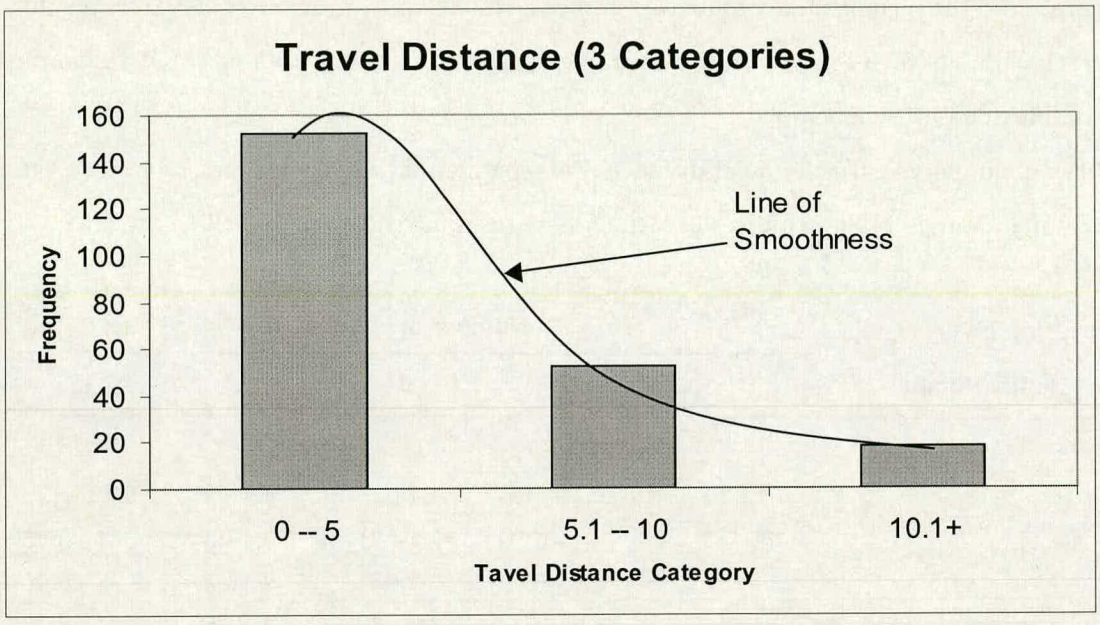


Figure 7.6a: Frequency graph of travel distance indexed into 3 indices

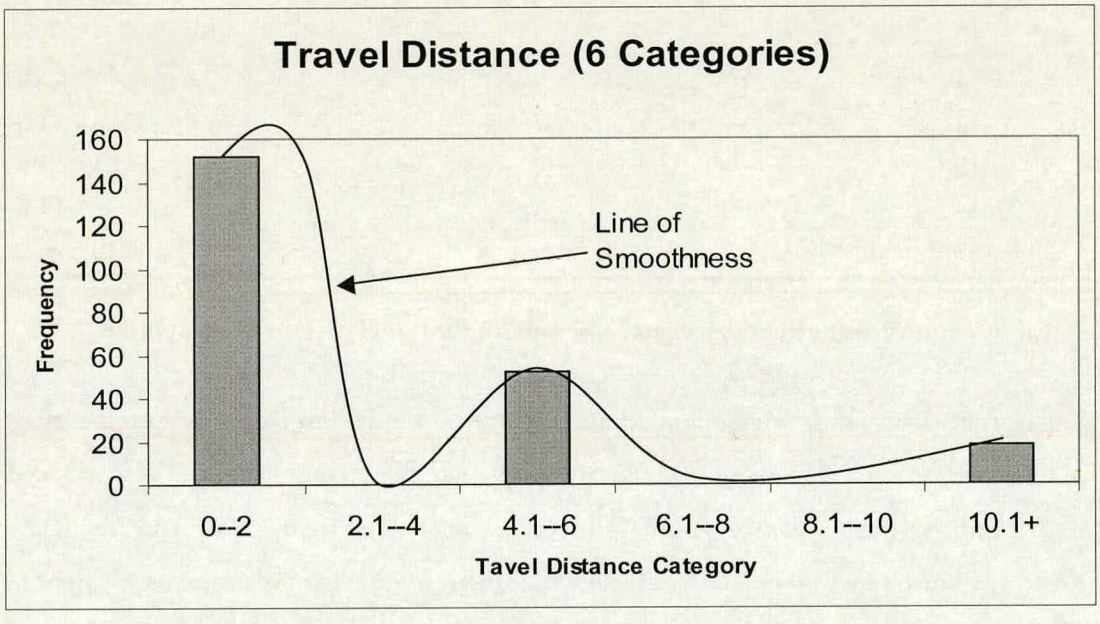


Figure 7.6b: Frequency graph of travel distance indexed into 6 indices

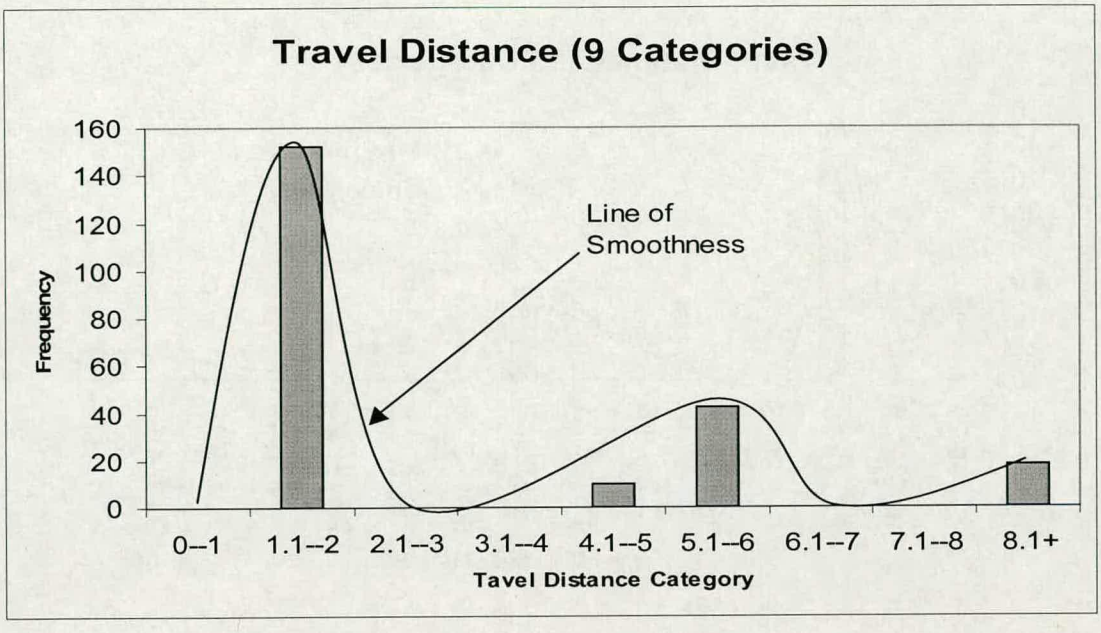


Figure 7.6c: Frequency graph of travel distance indexed into 9 indices

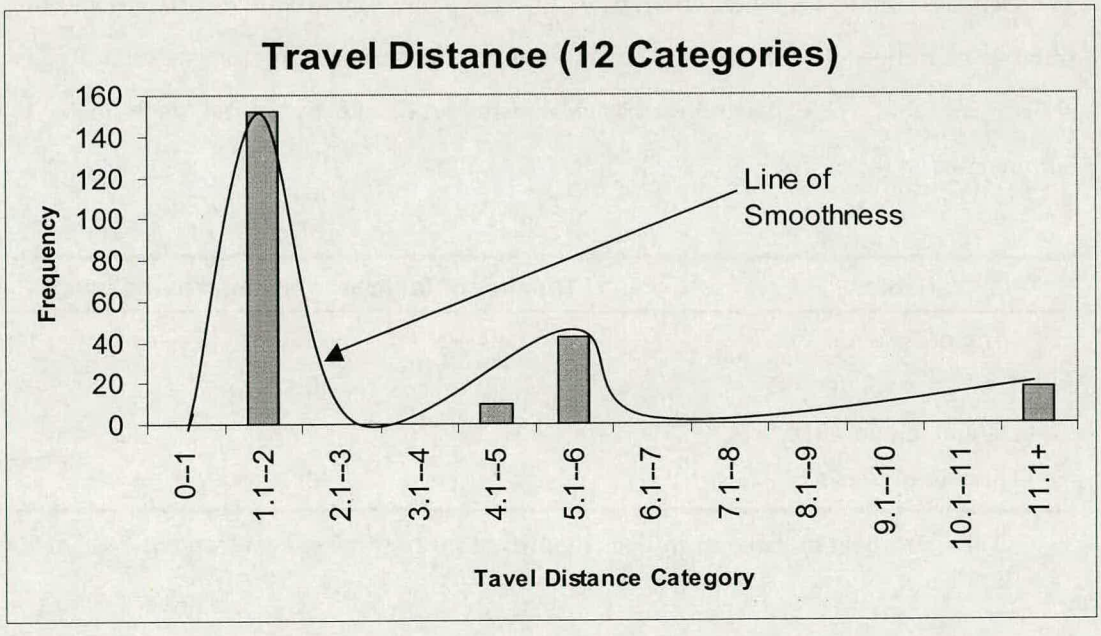


Figure 7.6d: Frequency graph of travel distance indexed into 12 indices

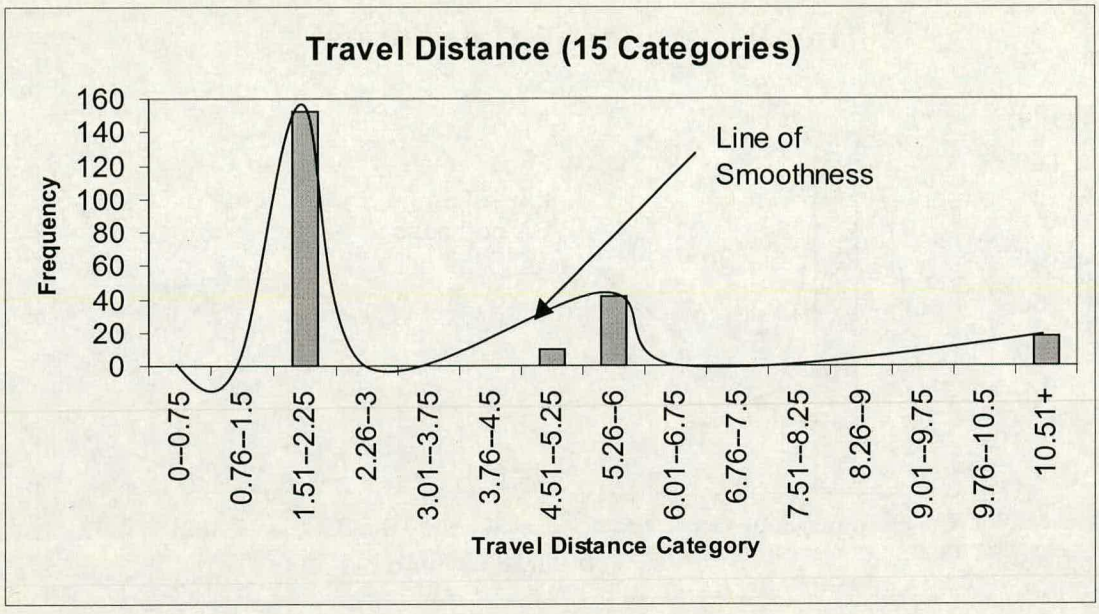


Figure 7.6e: Frequency graph of travel distance indexed into 15 indices

The visual assessment method was repeated for the 3 other quantitative variables and the number of indices that produced the smoothest frequency graphs for each variable is shown in table 7.2. The indexing mechanism produced by visual assessment is summarised in figure 7.7.

Variable	Number of Indices to be used in indexing
Travel distance	3
Total pour volume	6
Maximum temperature	6
Number of trucks	3

Table 7.2: Best number of indices identified through visual assessment

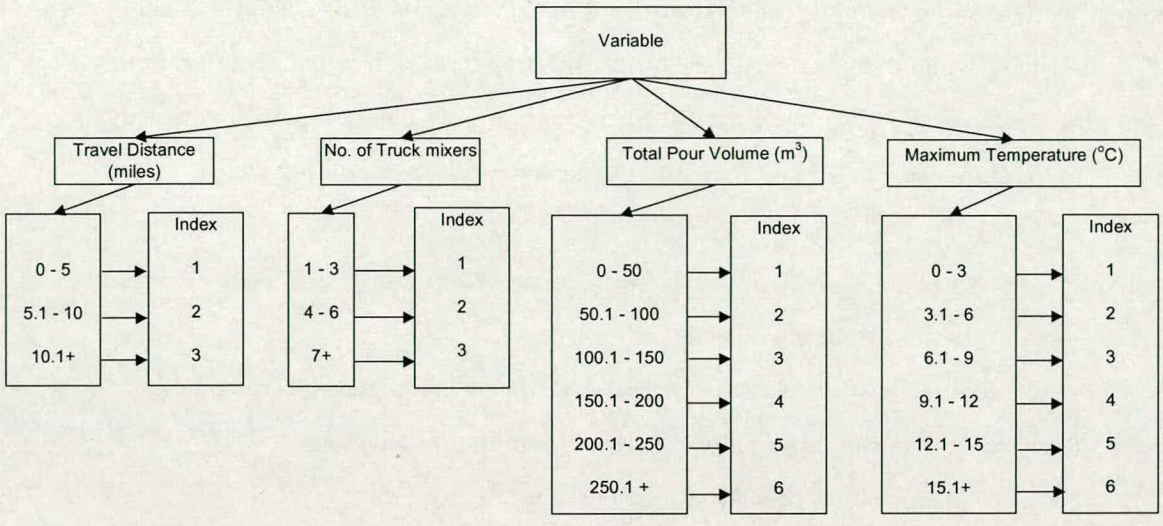


Figure 7.7: Crisp indexing mechanism defined by visual assessment

7.4.3 Fuzzy method

To restate, one of the difficulties in developing an indexing mechanism is in defining the number (or width) of indices in which to separate the quantitative variables. Above, 3 methods for defining crisp indices were presented. One of these indexing methods may well be the best method to use in CBR models of the RMC supply process. However, a distinction is being made within the data that, for example, a total pour volume of 79.3 m^3 is wholly different from a total pour volume of 79.4 m^3 when using Sturge's rule to define the number of indices to use. This distinction is clearly difficult to justify in the real world, but it is necessary when using crisp indexing methods. In order to make CBR more realistic this project has included a fuzzy indexing method, which avoids the need to index a variable in such a clear-cut manner as to say a total pour volume of 79.3 m^3 belongs in one set, while a total pour volume of 79.4 m^3 belongs in another.

Fuzzy indexing like crisp is governed by membership functions, which define where one index ends and another begins. The difference between fuzzy and crisp is that in the crisp case the membership functions cannot overlap, but in fuzzy the edges of indices are effectively blurred, allowing a variable's value to be partially a member of one index, while also being partial members of another(s); alternatively this can be described as membership overlap. To illustrate this concept of partial membership, return to figure 7.3 and equation 7.1 which

defined the crisp membership functions in the age classification example in section 7.4.1. These crisp membership functions would result in a person aged 64 and 363 days as being classified as *young*, while a person aged 65 days as being classified as *old*. This example highlights the problem of indexing data in a crisp manner: how could a person suddenly become old in the course of 2 to 3 days? In reality the person probably wouldn't.

Alternatively ages could be classified using fuzzy membership functions, such as those defined in figure 7.8. In this instance if a person is:

- <40 years old, they are fully classified or indexed (membership grade = 1) as *young*.
- >= 40 and <50, the person is partially a member of (or indexed as) *young*.
- >=50 and <60, the person is partially a member of *young* and *old*.
- >=60, the person is partially a member of *old*.
- >=70, the person is fully indexed as *old*.

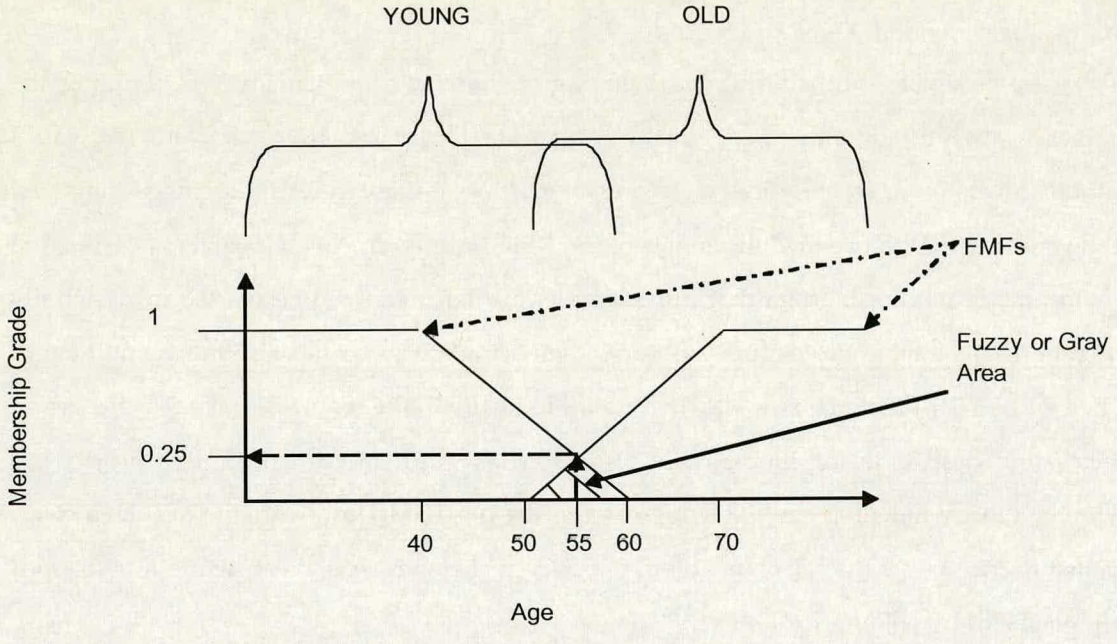


Figure 7.8: Fuzzy method of classifying ages in the example

It is important to define fuzzy membership functions adequately, as they govern the fuzzy indexing mechanism. A simple, iterative method of defining the fuzzy membership functions, and hence the size of the fuzzy indices is presented in the next section.

7.4.4 Defining the fuzzy index sizes (or fuzzy membership functions)

As mentioned in the previous section a simple iterative method has been used in this project to define the appropriate ranges of fuzzy membership functions, and hence the fuzzy index sizes. The methodology of the iterative method is shown in figure 7.9. To begin, decisions must be made as to the number and range of fuzzy membership functions for each variable that is to be indexed. These decisions were made and the resulting fuzzy membership functions are shown in figures 7.10a to 7.10d. Subsequently using the fuzzy indices derived from the fuzzy membership functions a model is developed and validated. The fuzzy indices that were derived from the fuzzy membership functions in figures 7.10a to 7.10d are shown in figure 7.11. Fuzzy indices differ from crisp because they allow partial membership to one or more indices. For example, in figure 7.11 if the travel distance recorded in a past operation is less than 2 miles, then that value is separated into index 1 with a membership grade of 1 (or full membership). Whereas if this travel distance were 4.5 miles, then the travel distance would be separated into indices 1 and 2 with a membership grade of >0 and <1 (partial membership).

Validation in this instance involves recreating the 50 past operations observed at the construction of the Thelwall Viaduct (see 4.5) and measuring the difference between the observed productivity and the model estimate of productivity in a correlated inspection approach to validation; this difference is known as the validation error. That is the completion of the first model iteration.

The subsequent iterations involve attempting to optimise the range of the fuzzy membership functions for each variable. In this instance the chosen fuzzy membership functions ranges were adjusted by +10%, +20%, +30%, +40%, -10%, -20%, -30%, and -40%. This resulted in a possible 9 (8 listed above, plus the chosen range) different ranges of fuzzy membership function. As there are four variables which are quantitative and require to be indexed in this manner the calculation of the optimum combination of fuzzy membership functions range is a large one, resulting in 9^4 (no. ranges$^{\text{no. variables}}$) or 6561 models being developed and validated. Clearly this bid to find the optimum combination of fuzzy membership function range could prove to labour intensive. However, using a modern computer and automating the process of model development through computer programming resulted in the production of the 6561 models in less than 24 hours. It is acknowledged that even this automated method is time

consuming, but model developers could utilise optimisation techniques such as linear programming, or genetic algorithms to minimise the development and validation effort; the investigation of such methods have not been undertaken in this project.

The final stage of the iterative method is to compare the 6561 validation errors produced by the (6561) models and choose the model, and hence the fuzzy indexing mechanism, that produced the smallest validation error. In this study the fuzzy indexing mechanism which produced the smallest validation error, and consequently will be used further is shown in figure 7.12. In general this fuzzy indexing mechanism was comprised of indices that were reduced in size from those originally chosen; this could be viewed as making the indices increasingly crisp (less overlap of fuzzy membership functions). Specifically, the optimum combination was found by adjusting the fuzzy membership functions range as follows: for travel distance - reduction by 10%; for the number of truck mixers – no adjustment; for total pour volume – reduction by 10%; for maximum temperature – reduction by 10%.

In figure 7.12, it is noticeable for the variable total pour volume that the indices have not been strictly reduced by 10%. This is because a 10% reduction in the range of a fuzzy membership function would have left gaps in the indexing mechanism i.e. indices 1 and 2 would not only have no overlap, but they would not even begin and end at the same pour volume. Thus if a total pour volume were presented that corresponded with a gap, no indexing of that value would occur; this is unacceptable in a CBR model. This problem was solved by making two adjacent indices (at least) begin and end on the same value (of total pour volume in the example).

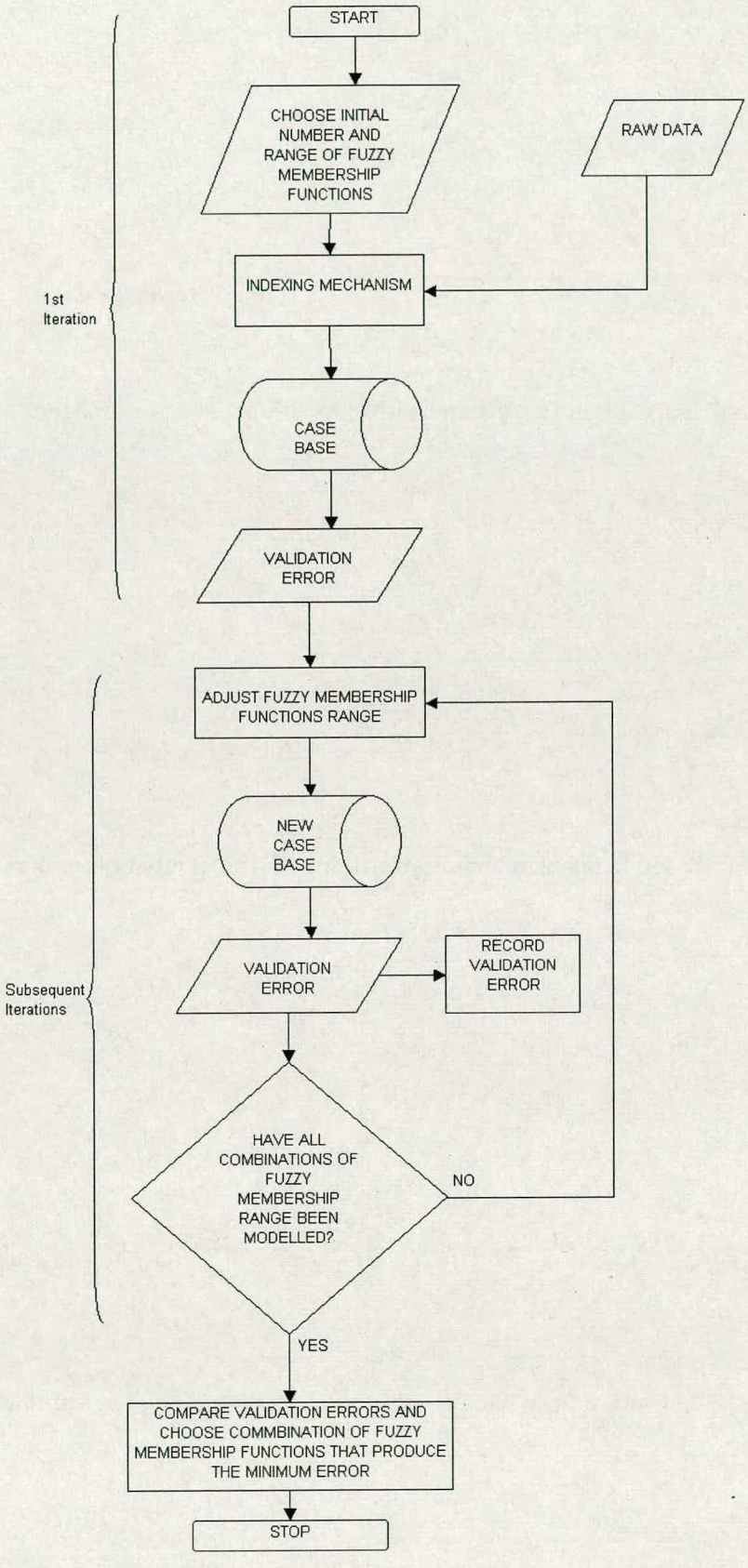


Figure 7.9: Methodology of the iterative definition of fuzzy membership functions

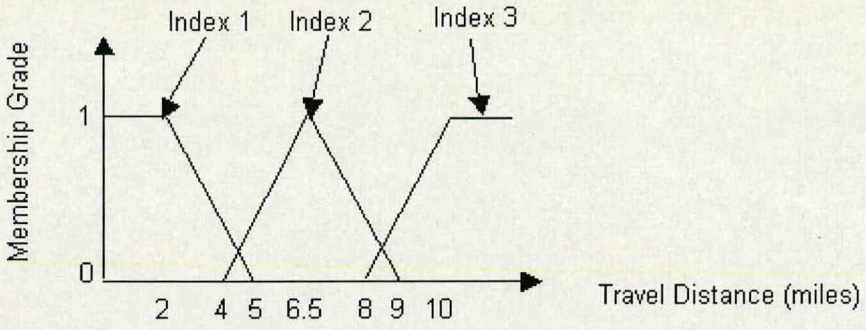


Figure 7.10a: Chosen fuzzy membership functions for travel distance

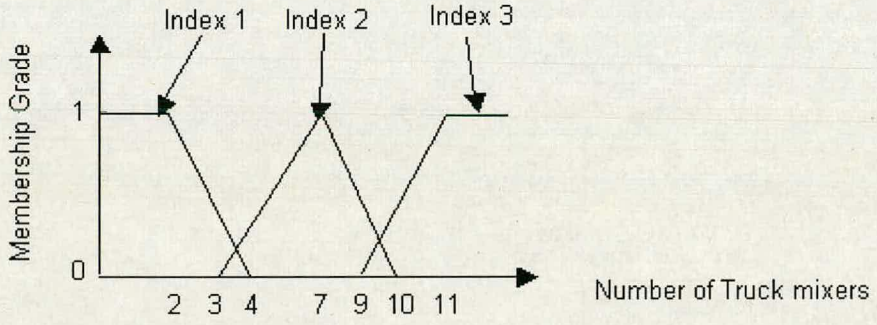


Figure 7.10b: Chosen fuzzy membership functions for the number of truck mixers

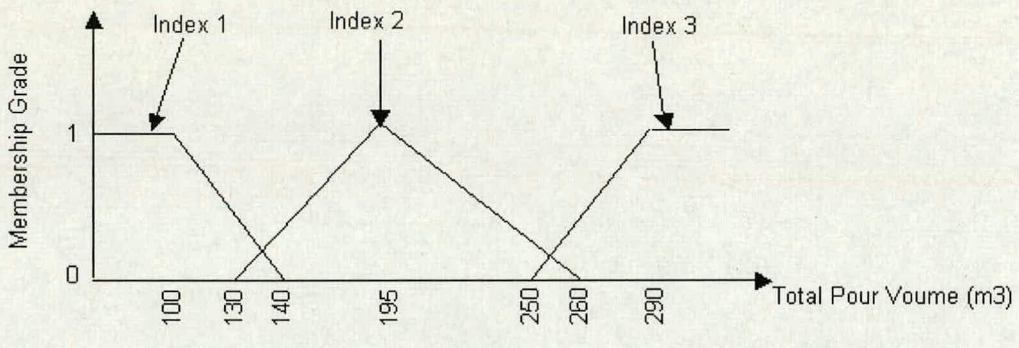


Figure 7.10c: Chosen fuzzy membership functions for total pour volume

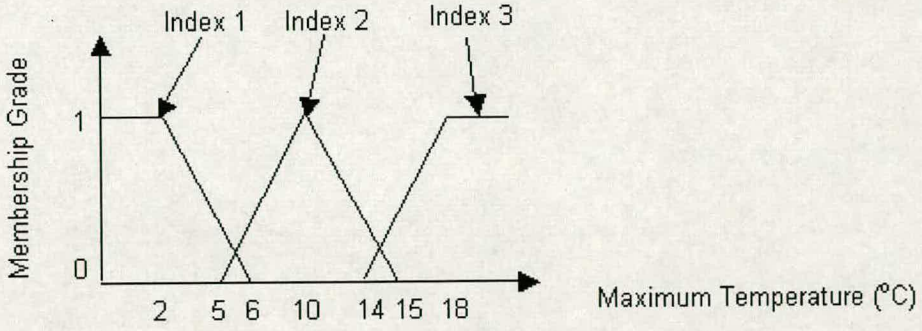


Figure 7.10d: Chosen fuzzy membership functions for maximum temperature

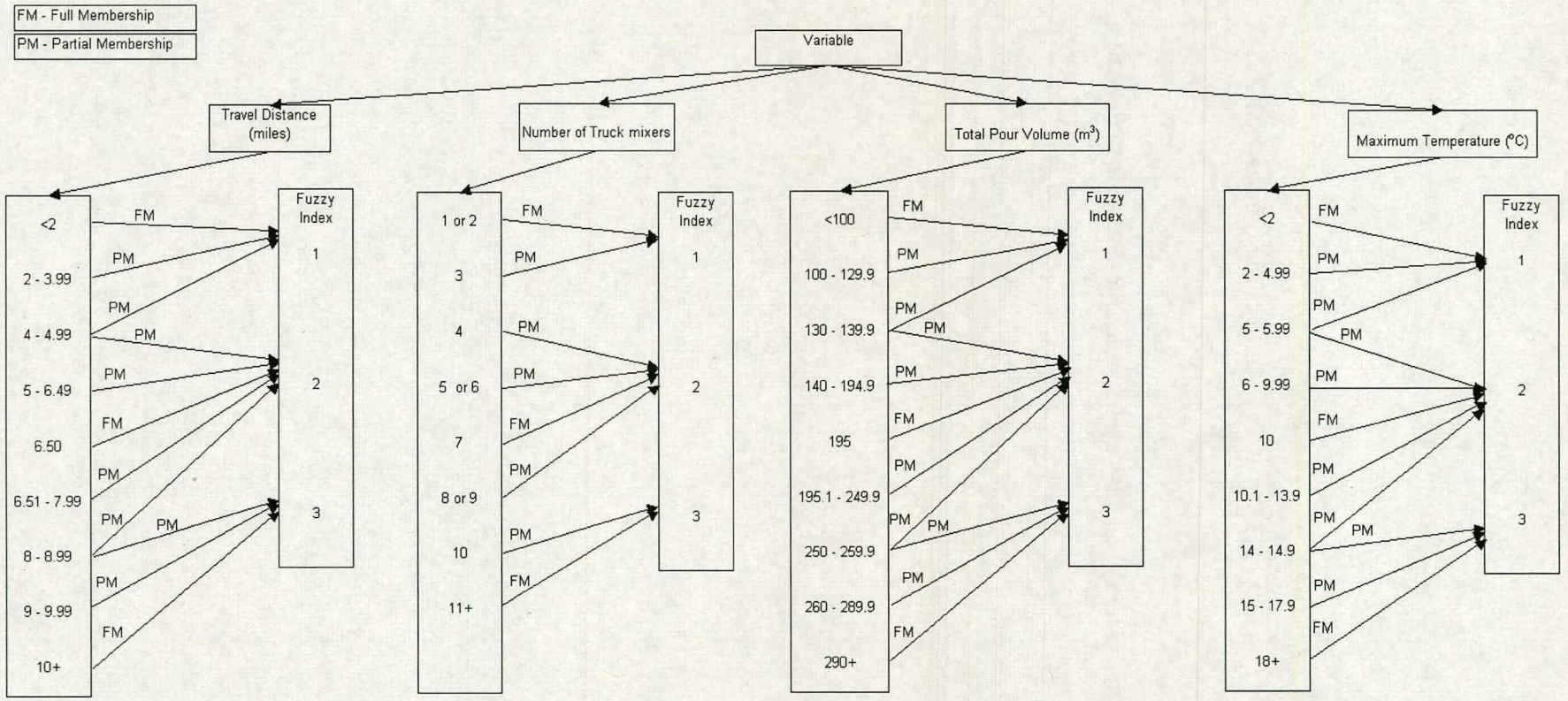


Figure 7.11: Fuzzy indexing mechanism produced by initial decision on fuzzy membership number and range

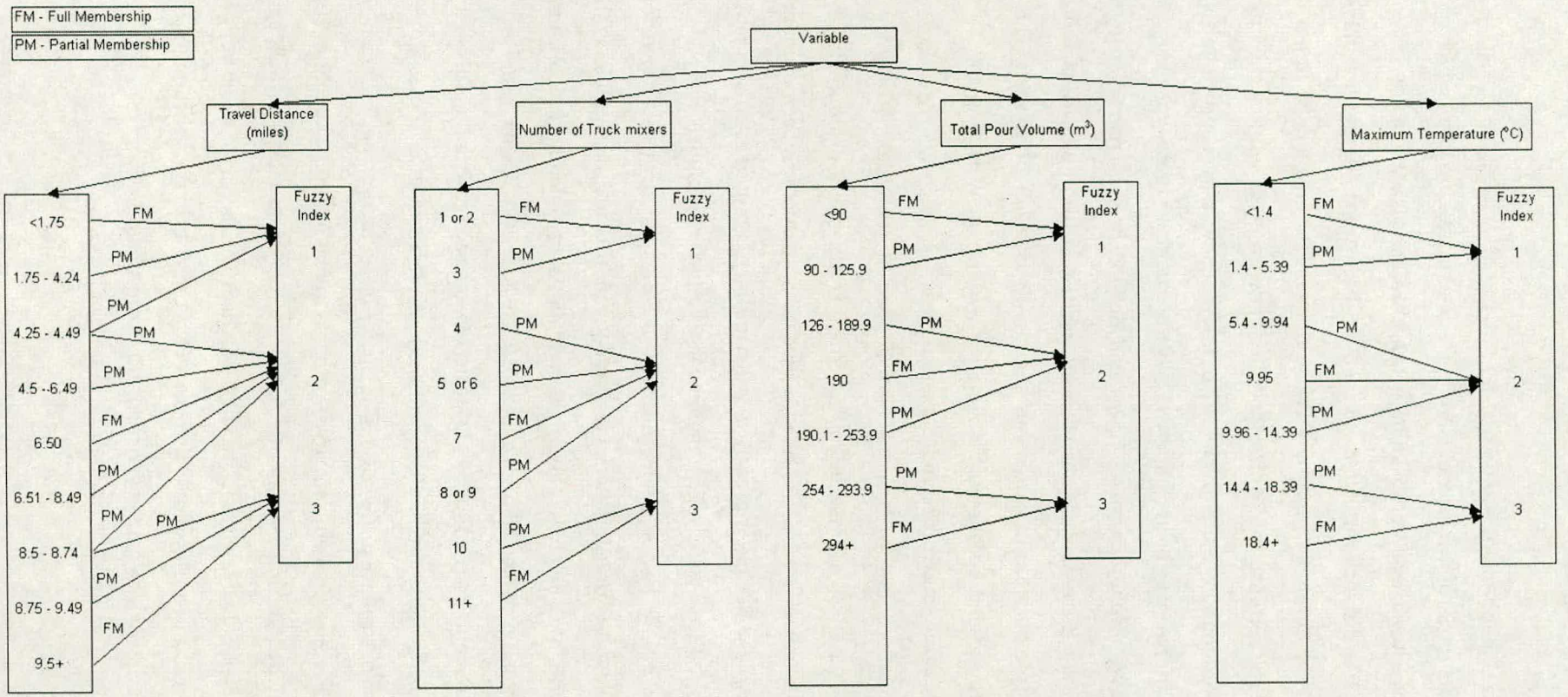


Figure 7.11: Final fuzzy indexing mechanism produced by the iterative method

7.5 Retrieval mechanism

The purpose of a retrieval mechanism is to find a case (in the case base) which can provide an acceptable answer to a current problem (target case) and this is usually done by calculating the similarity (using a distance measurement method) between the target and the stored cases. This purpose probably makes the retrieval mechanism the single-most important part of a CBR model. If the retrieval mechanism doesn't suggest a good solution then the model will produce a poor quality result. In a bid to find the most appropriate form of retrieval mechanism a number have been compared. These mechanisms are:

- **Crisp**. Where the case bases have been defined by the crisp indexing mechanisms (Sturge's, square-root and visual assessment) and the similarity measurements used are simple distance measurements commonly used in CBR models. This method has been considered because of its simplicity, which may help spread the use of CBR models in the construction industry.
- **Fuzzy**. Where the case base has been defined by the indexing mechanism shown in figure 7.12 and the similarity measurement is capable of dealing with measuring the distance between these "fuzzified" cases. This method has been examined because it possesses the flexible approach to data handling permitted by fuzzy methods, and this could produce a better model than the more rigid crisp models.
- **Cluster-based**. This is a novel method in CBR modelling developed for this thesis, which involves utilising the same technique used to identify outliers in the data set in chapter 5. This involves, clustering the data using a standard algorithm, finding the cluster upon which the search is to be focussed (using the Penrose distance measurement), measuring the distance between each case in this cluster and the target case (using the same similarity measurements used in crisp retrieval) to find the nearest (in distance terms) case. The cluster-based method has been considered in this project because it is efficient. In the other retrieval methods, the entire case base must be searched before a solution is found. Here, only a cluster (which can sometimes consists of only a few cases) must be searched. The increased efficiency of this method could result in a CBR model that is more acceptable to construction practitioners, due to the time savings made over crisp and fuzzy methods.

As mentioned above all retrieval mechanisms involve making a similarity measurement between a target case and the stored cases, to find the most similar. Regardless of which mechanism is being considered here, there are two types of similarity measurement to be made: non-weighted and weighted. A non-weighted method gives equal importance to all of the variables in a measurement of similarity (Hand et al., 2001). The benefit of this is that the relative importance of variables does not have to be defined. This could result in a model that is less realistic and thus models in which the variables have their importance assigned to them have also been considered. Linear regression analysis is used to calculate variable weighting and the results are presented in the next section. Variable weighting affects the ability of a retrieval mechanism to suggest an appropriate solution to a problem. In section 7.5.2, an investigation has been undertaken to determine how a change in variable weighting affects the retrieval mechanism – if a change in weighting is significant a number of weighting calculations may need to be used to maximise the probability of producing an accurate retrieval mechanism – if a change in weighting is not significant, the method mentioned above should be sufficient. Descriptions of the crisp, fuzzy and cluster-based retrieval mechanisms are presented in 7.5.3, 7.5.4 and 7.5.5, respectively.

7.5.1 Variable weighting calculation

Linear regression was used in this project to determine the key variables of the process (see 5.6.2). The outcome of linear regression is of the form:

y = mx + c, where y is an output, m is a weighting, x is an explanatory variable and c is a constant defined by the least squares (LS) mathematical function. The linear regression presented in 5.6.3 resulted in the following equation:

$$\begin{aligned} y = & -0.5455 \times \textit{Month of Pour} - 0.0645 \times \textit{Travel Dis}\,\tan ce - \\ & 1.6783 \times \textit{Type of Pour} + 0.9696 \times \textit{Number of Truckmixer s} \\ & + 0.1603 \times \textit{Total Pour Volume} - 0.0478 \times \textit{Maximum Temperatur e} \\ & + 22.4788 \end{aligned} \tag{7.4}$$

The absolute values of the weightings in this equation can be extracted and normalised so that all the weightings in the equation combine to equal 1. This normalisation has been done and the results are shown in table 7.3.

Variable		Weighting
Month of Pour		0.16
Travel Distance		0.02
Type of Pour		0.48
Number of Truck mixers		0.28
Total Pour Volume		0.05
Maximum Temperature		0.01
	Cumulative	1

Table 7.3: Variable weightings calculated using linear regression

7.5.2 Variable weighting sensitivity analysis

A sensitivity analysis was performed to determine the effect on the retrieval mechanism of a change in a variable's original weighting by: -50%, -20%, -10%, +10%, +20%, and +50%. This analysis was done one variable at a time, while the others remained constant. The analysis involved using each model produced by the sensitivity analysis to retrieve the five best solutions to the same case and to record the case number of these solutions. If there is a change in the cases proposed as the best solution due to changes in a variable's weighting then the retrieval mechanism can be declared sensitive to these changes.

The results of the sensitivity analysis are shown in table 7.4 and it can be concluded that small changes (up to +/- 20%) in a variable's weighting had negligible effect on the retrieval mechanism. However, a change of +/-50% in the weightings of month of pour, type of pour, number of truck mixers and total pour volume resulted in different cases being proposed (no. 63 instead of no. 11) as the best solutions (ranked 1st in table 7.4) from those when the weightings were originally set by linear regression. To investigate the effect of this change the productivities of the target case was compared with those of cases 11 and 63 (see figure 7.13). Clearly there is no significant difference between the solutions provided by either case 11 or case 63 – they are both relatively accurate – and it can be concluded that the retrieval

mechanism is not sensitive enough to changes in the weightings of variables to warrant the use of alternative methods to calculate these weightings.

		Percentage difference from original weighting						
Variable	**Ranking**	**-50**	**-20**	**-10**	**0**	**10**	**20**	**50**
Month of Pour	1	63	11	11	11	11	11	11
	2	11	63	63	63	63	63	63
	3	25	25	25	25	25	25	25
	4	33	33	33	33	33	33	33
	5	85	85	85	85	85	85	85
Travel Distance	1	11	11	11	11	11	11	11
	2	63	63	63	63	63	63	63
	3	25	25	25	25	25	25	25
	4	85	85	85	85	85	85	85
	5	33	33	33	33	33	33	33
Type of Pour	1	11	11	11	11	11	11	63
	2	63	63	63	63	63	63	11
	3	178	25	25	25	25	25	25
	4	25	178	33	33	33	33	33
	5	85	33	85	85	85	85	85
Number of Truck mixers	1	11	11	11	11	11	11	63
	2	63	63	63	63	63	63	11
	3	85	17	17	17	17	17	17
	4	17	85	33	33	33	33	178
	5	33	33	85	85	178	178	33
Total Pour Volume	1	11	11	11	11	11	63	63
	2	63	63	63	63	63	11	11
	3	85	85	25	25	25	25	25
	4	13	25	85	33	33	33	33
	5	33	33	33	85	178	178	178
Maximum Temperature	1	11	11	11	11	11	11	11
	2	63	63	63	63	63	63	63
	3	25	25	25	25	25	25	85
	4	85	33	33	33	33	33	25
	5	33	85	85	85	85	85	33

Table 7.4: Top five ranked cases in weightings sensitivity analysis

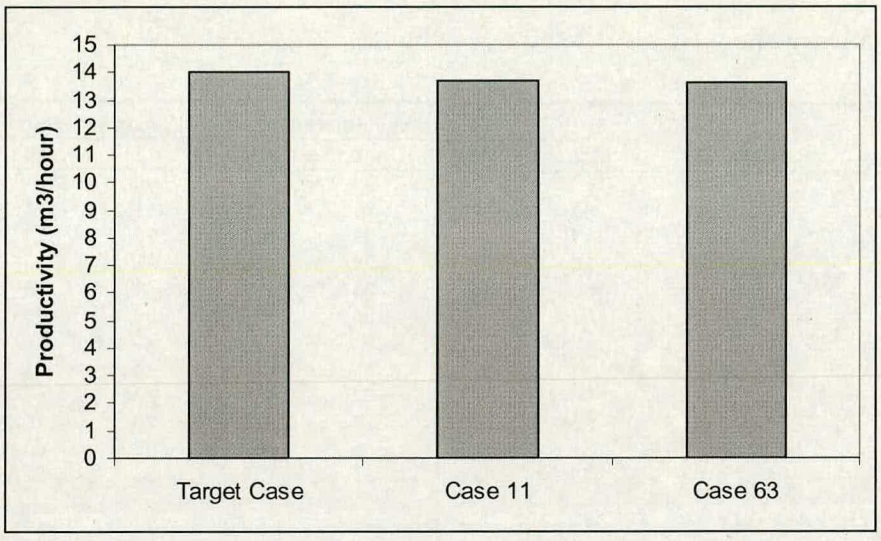


Figure 7.13: Comparison of productivities in weightings sensitivity analysis

7.5.3 *Crisp retrieval mechanism*

A crisp retrieval mechanism is based upon making simple measurements of the similarity between a (crisp indexed) target case and each of the stored cases, where all of these cases have been separated into crisp indices, except the qualitative variables. This measurement of similarity is made in two stages:

- First, a measure of the similarity between the same variables in the target and stored cases is made. This is done using the rule:

$$sim_i(i_{Target}, i_{Stored}) = \begin{cases} 1 & if\ i_{Target} = i_{Stored} \\ 0 & otherwise \end{cases} \quad (7.5)$$

  where i_{Target} is a variable from the target case; i_{Stored} is a variable from a stored case; $sim_i(i_{Target}, i_{Stored})$ is measurement of similarity between two variables.

- Then a measurement of the similarity between two cases (target and stored) is made using the calculation:

$$sim(Target, Stored) = \frac{1}{p}\sum_{i=1}^{p} sim_i(i_{Target}, i_{Stored}) \quad (7.6)$$

  where sim(Target, Stored) is a measure of the similarity between a stored and target case; and p is the number of variables.

The above method has not included the relative importance (weightings) of variables. To do so results in a minor change in equation 7.6 to:

$$weighted\ sim(T\arg et, Stored) = \frac{1}{p}\sum_{i=1}^{p} w_i sim_i(i_{T\arg et}, i_{Stored}) \quad (7.7)$$

where weighted sim(Target, Stored) is a measure of the similarity between a stored and target case taking into account the relative importance of variables; w_i is the weighting of a particular variable, i.

The calculations (equations 7.6 and 7.7) deal with both qualitative and quantitative variables, as they are making a comparison between two values on a like-for-like basis, e.g. if the target case has a base pour and a stored case has a base pour then the $sim_i(i_{Target}, i_{Stored}) = 1$, where i is the variable type of pour.

After a measurement of similarity is made between the target case and each case in the case base, the most similar case is retrieved and provided to the learning mechanism (discussed in 7.2), which takes the output from the retrieved case and matches this output with the inputs (6 variables) from the target case, to form a new case. This case is both presented as a solution and stored in the case base as a new case.

7.5.4 *Fuzzy retrieval mechanism*

The quantitative variables of the target case are indexed using the fuzzy indexing mechanism presented in figure 7.12. These indexed variables are combined with the qualitative variables to form a fuzzy target case.

The similarity between the fuzzy target case and each case stored in a case base is then calculated. As in the crisp retrieval mechanism this calculation of similarity is performed in two stages by first calculating the similarity between the variables of the cases in question (target and stored) and then calculating the similarity between the cases themselves. Here similarity between the variables of the target and stored cases is of the form:

$$d_i = \sum\nolimits_j abs(x_{ijk} - x_{ijn}) \quad (7.8)$$

Where x_{ijk} and x_{ijn} are the membership indices of variable i (for example, travel distance may be partially a member of index 1 and index 2), with fuzzy membership grade, j (e.g. a fuzzy

membership grade of 0.8 to index 1, and a fuzzy membership grade of 0.4 to index 2) for cases k (stored case) or n (target case).

The overall similarity between cases, d is defined by:

$$d = \sum d_i \tag{7.9}$$

The above calculations do not take into account the weightings of variables. To do so involves a minor change to equation 7.9, so that is of the form:

$$d = \sum w_i d_i \tag{7.10}$$

Where w_i is the weighting of a particular variable, i.

The most similar case has the shortest distance measurement, i.e. smallest d and it is this case that is used by the learning mechanism; As in the crisp case, the output from the retrieved case is matched with the inputs (6 variables) from the target case, to form a new case. This case is presented as a solution and stored in the case base as a new case.

7.5.5 *Cluster-based retrieval mechanism*

This cluster-based retrieval mechanism is novel to CBR model development and was devised and developed in this research project. It is based upon the data clustering method presented in chapter 5, which was used to identify outliers in the data.

The cluster-based retrieval mechanism does not require the cases (either target or stored) to be indexed in any way. This may seem to be removing the efficiencies (measured in seconds) in the retrieval mechanism brought about by reducing the range of a variables value from say roughly 270m³ in the case of pour volume to 14, when indexing was performed using the square-root rule. However, this cluster-based retrieval mechanism provides efficiency in another way – as the search for a solution is quickly narrowed down to the nearest cluster to the target case – and this simple step results in a higher retrieval efficiency than that witnessed in a crisp or fuzzy indexing mechanism. For instance a typical retrieval time for either a crisp or fuzzy CBR model is 600 seconds, while the maximum recorded retrieval in a cluster-based model has been 120 seconds. This saving of at least 8 minutes could make a large difference in the acceptance of the technology among construction practitioners.

The steps involved in the cluster-based retrieval mechanism may appear to be cumbersome, but each step requires only a small amount of computing power and they are all automated. These steps are:

- Perform a k-means cluster analysis of the cases stored in the case base, where k =3. This results in the cases being separated into 3 clusters and a mean value of each variable in each of the 3 clusters being calculated.
- Calculate the covariance matrix of each variable held in the cases in the case base. This is performed using a standard function in Microsoft Excel.
- Calculate the Penrose distance between the target case and the central-position (centre) of each of the 3 clusters. This calculation is of the form:

$$P_{Target,Cluster\,1} = \frac{(Month\ of\ Pour_{Target} - \mu_{Month\ of\ Pour,Cluster\,1})^2}{P \times V_{Month\ of\ Pour}} + ... + \frac{(Maximum\ Temperatur\ e_{Target} - \mu_{Maximum\ Temperatur\ e,Cluster\,1})^2}{P \times V_{Maximum\ Temperatur\ e}}$$

(7.11)

Where: $P_{Target,\ Cluster\ 1}$ is the Penrose distance between the target case and cluster 1; $Month\ of\ Pour_{Target}$ is the value of month of pour in the target case; $\mu_{Maximum\ Temperature,\ Cluster\ 1}$ is the mean value for the variable maximum temperature of all of the cases contained in cluster 1; P is the number of variables; $V_{Month\ of\ Pour}$ is the covariance matrix of the variable month of pour (calculated in step 2).

- The cluster which corresponds to the minimum Penrose distance calculated in the previous step is the one on which focus is now placed; this cluster is closer to the target case than the other clusters. This is the step that produces the efficient retrieval of cases that was discussed above.
- Form a new case base that contains only the cases that formed the cluster identified in the previous step.
- Perform a measurement of similarity between each case in the new, reduced case base and the target case. This measurement is borrowed from the crisp retrieval mechanism. Hence when no variable weightings are being considered similarity is calculated using the formulae of equation 7.12, followed by equation 7.6 in the manner described in 7.5.3. When weightings are being included in the similarity calculation the formulae of equations 7.5 and 7.7 are used instead.

$$sim_i(i_{Target}, i_{Stored}) = \begin{cases} 1 & if\ i_{Target} = i_{Stored \pm 10\%} \\ 0 & otherwise \end{cases} \quad (7.12)$$

where i_{Target} is a variable from the target case; i_{Stored} is a variable from a stored case; $sim_i(i_{Target}, i_{Stored})$ is measurement of similarity between two variables.

After a measurement of similarity is made between the target case and each case in the case base, the most similar case is retrieved and provided to the learning mechanism (discussed in 7.2), which takes the output from the retrieved case and matches this output with the inputs (6 variables) from the target case, to form a new case. This case is both presented as a solution and stored in the case base as a new case.

7.6 CBR model development and validation

A number of CBR models were developed comprising of various combinations of the indexing and retrieval mechanisms presented in 7.4 and 7.5, in an attempt to find the combination that produced the most accurate model. To determine this accuracy each model was validated. The most accurate model composition will subsequently be used in the hybrid model, *CBRSim*.

Model Number	Indexing Mechanism	Retrieval Mechanism
1	Crisp (Sturge's)	Crisp (Non Weighted)
2	Crisp (Sturge's)	Crisp (Weighted)
3	Crisp (Square-root)	Crisp (Non Weighted)
4	Crisp (Square-root)	Crisp (Weighted)
5	Crisp (Visual assessment)	Crisp (Non Weighted)
6	Crisp (Visual assessment)	Crisp (Weighted)
7	Fuzzy (Iterative)	Fuzzy (Non Weighted)
8	Fuzzy (Iterative)	Fuzzy (Weighted)
9	Raw Data (no indexing)	Cluster-Based (Non Weighted)
10	Raw Data (no indexing)	Cluster-Based (Weighted)

Table 7.5: Details of developed CBR models

7.6.1 CBR model development

The compositions (of indexing and retrieval mechanisms) of the models developed and validated in this study are shown in table 7.4.

The models of the composition shown in table 7.4 were developed and validated. The results of validation shall be discussed in the next section.

7.6.2 CBR model validation

The process of model validation was to use the data that was reserved during the development process – 20% of the original data set, or 46 cases – as target cases and provide this validation data set to each model. A productivity estimate was produced for each target case by each model and this estimate was then directly compared with the observed productivity (as recorded in each target case) in a correlated inspection approach (see 6.6). The difference between estimated and observed is known as the validation error (in percentage terms). An average of the validation error was calculated over the whole validation data set (of 46 cases), and the average validation error of each model can be compared to determine the most accurate indexing and retrieval combination to use in the hybrid model, *CBRSim*.

The results of this validation exercise are shown in table 7.5. It is interesting to note that different sizes of indices produce models of differing capabilities. For instance, the crisp models possess different degrees of accuracy, with the Sturge's rule-defined indices performing best.

The fuzzy indexing and retrieval models were more accurate than two of the crisp models and more research should be undertaken to examine alternative methods of defining the sizes of fuzzy indices and the effect that this has on model estimation accuracy.

It is noticeable when comparing a model that has included the weightings of variables with the same model that has not, the weighted model is always more accurate in this study. Including the weightings of variables already has the advantage that a more realistic model is being produced (than a non-weighted one) and this validation exercise has further justified their inclusion.

Clearly, model 10 is the most accurate. This model comprises a cluster-based retrieval mechanism and takes into account the weightings of variables and as the most accurate it shall be used in the CBR part of the model, *CBRSim*, which shall be presented in the next chapter.

Model Number	Indexing Mechanism	Retrieval Mechanism	Average Validation Error (%)
10	Raw Data (no indexing)	Cluster-Based (Weighted)	7.3
9	Raw Data (no indexing)	Cluster-Based (Non Weighted)	8.4
2	Crisp (Sturge's)	Crisp (Weighted)	11.7
1	Crisp (Sturge's)	Crisp (Non Weighted)	12.2
8	Fuzzy (Iterative)	Fuzzy (Weighted)	14.9
4	Crisp (Square-root)	Crisp (Weighted)	15.3
7	Fuzzy (Iterative)	Fuzzy (Non Weighted)	15.7
3	Crisp (Square-root)	Crisp (Non Weighted)	15.8
6	Crisp (Visual assessment)	Crisp (Weighted)	16.5
5	Crisp (Visual assessment)	Crisp (Non Weighted)	18.6

Table 7.6: CBR model validation results

7.7 Chapter Summary and Conclusions

A simulation model was developed (and presented in chapter 6) and declared to be too complex to be practical because a user is required to choose the probability distributions and parameters to drive the model. This was dubbed the complexity problem and case-based reasoning (CBR) was investigated as a means of solving this. A CBR model works by using past cases that are held in a database (case base) to suggest a solution to a new (target) problem. A typical CBR model comprises: an indexing mechanism – to store cases in the case base in a standard form; retrieval mechanism – to find similar cases to the target; adaptation mechanism – to adapt the retrieved case to closer match the target; and a learning mechanism – to store the target case and its solution as a new case for future use. The development of these mechanisms is a heuristic process, which was undertaken in this research. To summarise by mechanism:

Indexing mechanism. Most CBR models make the assumption that the data being dealt with is of a qualitative type, and there has been little discussion as to how quantitative-type data

should be indexed and retrieved. This is a problem in modelling cyclical and stochastic construction processes where the majority of the observed data is quantitative. To overcome this problem, a method was used that is common in the formulation of histograms – to separate the values of a variable into separate categories (or indices). Two types of category (index) were considered: crisp, where a variable's value is either a member of an index or not; and fuzzy, where a variable's value can hold some membership to multiple indices.

Three methods of defining crisp index size's were investigated – Sturge's rule, square-root rule and visual assessment. Where visual assessment involved choosing a number of indices in which to separate (or index) the quantitative variables, plotting the frequency graph produced from the data set and finally selecting the number of indices that produces the smoothest frequency graph.

The fuzzy indices were defined using an iterative method, involving a subjective decision on the number of and range of fuzzy membership functions (fuzzy indices) for each variable. Then the chosen fuzzy membership functions were adjusted by +/- 40% in increments of 10%. This resulted in 9 different variations in fuzzy membership function. Given that there were 4 variables being indexed, to fully investigate all possible combinations a total of 6561 models were produced (9^4 models or $no.ranges^{no.variables}$). Finally, the validation errors from the 6561 models were compared and the model that produced the lowest error was selected for further use.

It was found that the fuzzy membership functions that produced the best model were all smaller than the original (subjectively defined) membership functions. This points to either the subjective definition being poorly made (ranges were too large) or crisp indices provide more accurate models than fuzzy, as the ranges are moving toward crisp ranges (there is less overlap in the range).

Retrieval mechanism. The purpose of a retrieval mechanism is to find a case that can provide an acceptable answer to a current problem. The process of finding a case is usually performed by measuring the similarity (with a mathematical distance measurement) between the current case and all the cases stored in a data (case) base. Three types of retrieval mechanism were investigated: crisp; fuzzy; and a novel cluster-based method.

Regardless of which of the above mechanisms were being used there were two forms of distance measurement involved: weighted and non-weighted. A weighted measurement involved applying weightings of importance to the variables – in a bid to make the model more realistic. A sensitivity analysis to changes in the weightings was performed to examine the impact of assigning poorly these weightings. It was found that the model's retrieval capabilities were wholly unaffected by small changes in the weightings (up to +/- 20%), but were affected by larger changes (+/- 50%). This means that a model was capable of producing a good solution to a problem even when the changes made to weightings were up to +/20 %. Further, a change of +/- 50% resulted in the model suggested a different solution, but this suggested answer also provided a good solution to the problem at hand. It may be concluded that the values of weightings are not as important as first suspected – as even significant changes in their value do not result in significant changes in the model's output.

The cluster-based method does not require the data to be indexed. It involves quickly rejecting cases that are unlikely to provide an answer to the problem and focusing the search on a now greatly reduced batch of potential solutions. This mechanism produces significant savings in retrieval time, as a crisp or fuzzy model takes 600 seconds to retrieve a solution, whereas the cluster-based method retrieves a solution in 120 seconds. This could make the model all the more attractive to construction practitioners.

Validation. In total 10 models were developed: 6 crisp indexed using the standard histogram formation methods (weighted and non-weighted); 2 fuzzy (weighted and non-weighted) and 2 models with cluster-based retrieval mechanisms (weighted and non-weighted).

These models were validated and the weighted cluster-based model performed best. This may be because: the retrieval mechanism rejects unlikely candidates from consideration, removing the likelihood of the model presenting a poor result; and the method does not require any indexing, thus avoiding any information loss that commonly occurs in the indexing process.

CHAPTER 8

THE HYBRID MODEL, *CBRSim*

Case-Based Reasoning (CBR) has now, clearly, been established as an effective method for choosing the correct distribution for simulation model input. This chapter takes this to its logical conclusion – by combining CBR with discrete-event simulation into a more effective hybrid.

8.1 Introduction

In chapters 6 and 7 the development and validation of a discrete-event simulation (DES) model of the RMC supply process, *ExcelSim*, was presented. Although the validation results proved promising, there was an overwhelming sense that this model had fallen into the same trap as all others preceding it. This trap was that the model was overly complex at the user-end due to the requirement that the user possess knowledge about the probability distributions that represent the durations of activities within the process. Without finding a solution to this problem, the model was in danger of becoming consigned to the annals of history as an academic exercise. This was not the aim of this PhD project (see 2.7); rather the aim was to develop a model capable of practical application that could be embraced by academics and industrialists alike. Hence, the focus of this project turned toward finding a method of removing the need for a user to possess statistical knowledge of the RMC supply process. The method that was considered was case-based reasoning (CBR), an artificial intelligence modelling paradigm (see chapter 7).

To recap, in chapter 7 ten models were developed. This was because the determination of the optimum combination of parts of a CBR model is a heuristic procedure. Each of these models were validated and it was determined that the model with no indexing mechanism – that is to say, the model uses raw data – and a weighted cluster-based retrieval mechanism was selected to be used in a hybrid model that selects the probability distributions that describe a process – removing this task from the construction practitioner. This hybrid model, *CBRSim* is presented and validated in this chapter, while the source code of this model may be found in Appendix D and a demonstration version of this model has been provided with this thesis. Subsequently, a comparison is made between *CBRSim*, *ExcelSim* (from chapter 6) and CBR model 10 (from chapter 7) to determine if there is any increase in model accuracy combined with an increase in the ease-of-use by forming a hybrid solution to the problem.

Chapter 8 is structured as follows:

- **Section Two.** The hybrid model, *CBRSim* is presented.
- **Section Three.** *CBRSim* is validated, followed by a comparison of the model with *ExcelSim* (from chapter 6) and CBR model 10 (from chapter 7).
- **Section Four.** The limitations of *CBRSim* are discussed.

- **Section Five.** An example of how *CBRSim* can be used by a construction practitioner to order RMC is detailed.
- **Section Six.** Chapter conclusions.

8.2 The hybrid model, *CBRSim*

The hybrid model, *CBRSim* consists of two modules (see figure 8.1):

- A cased-based reasoning (CBR) module.
- A discrete-event simulation (DES) module.

8.2.1 CBR Module

The CBR module of *CBRSim* holds the same indexing and retrieval mechanisms as CBR model 10 from chapter 7 as these mechanisms were the most accurate and efficient of the options considered in chapter 7. To reaffirm, CBR model 10 had no indexing mechanism per se and that is because its (cluster-based) retrieval mechanism works on raw data, resulting in no requirement for case indexing to take place.

The case base is different in the hybrid model (from the CBR models of chapter 7). To recap, the models of chapter 7 had a case base consisting of: six input variables - month of pour, type of pour, travel distance, number of truck mixers, pour volume, and maximum temperature; and four outputs – productivity, cost, utilisation of truck mixers, and utilisation of on-site resources.

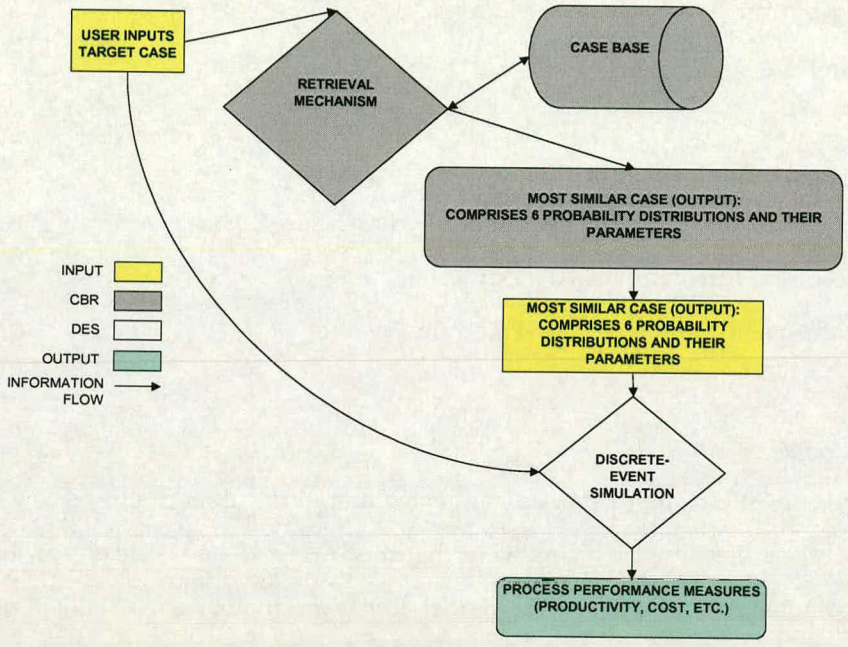


Figure 8.1: Methodology of *CBRSim* model

In the CBR module of *CBRSim*, the input variables held in the case base remain the same - it is the outputs which differ. This is because the objective of CBR is not to produce a final result (e.g. productivity estimates). Rather the aim is to flow information, in the form of a probability distribution (and its parameters) that represent the duration of each activity in the RMC supply process (see figure 8.2 for these activities), to a discrete-event simulation model that will produce a final result (process performance estimates).

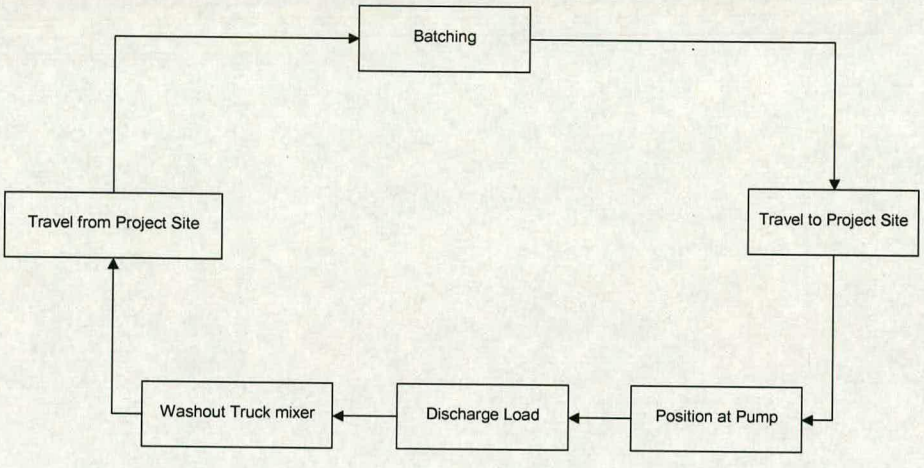


Figure 8.2: The activities within the RMC supply process

A sample of the case base is shown in figure 8.3 and it is noticeable that there are many different varieties of probability distribution (e.g. beta or lognormal). For the duration of each activity (e.g. batch time) in a particular case, 27 different probability distributions have been fitted to the observed data (of that particular case) and the parameters of these distributions have been calculated. The process of fitting these 27 probability distributions to the activity durations of the 232 observed cases was detailed in chapter 6. To recap, it was found that only 7 out of the 27 distributions provided an adequate fit to the entire data set. The 7 distributions were: uniform, normal, lognormal, exponential, weibull, gamma and beta. Thus in the case base of *CBRSim*, the output of a case is made up one of the 7 distributions listed above (and parameters) representing each activity duration.

The CBR module provides a graphical user interface (see figure 8.4) that allows a user to provide an input to the *CBRSim* model. The information required is a value for each of the 6 input variables and this information forms a target case to be solved.

When the appropriate information has been supplied by a construction practitioner the CBR module utilises its indexing and retrieval mechanisms to search through all of the cases held in the case base and find a case which has similar details to those of the target case.

Once the CBR module has produced an output, its role is at an end (see figure 8.1). This output, in the form of the probability distributions (and parameters) that represent the 6 activities within the RMC supply process, then becomes part of the input to the discrete-event simulation module.

Case Number	Month of Pour	Type of Pour	Travel Distance (Miles)	Number of Trucks	Pour Volume (cum)	Maxmium Temperature (oC)	Batch Time (minutes)			Travel to Site Time (minutes)			Position Time (minutes)			Discharge Time (minutes)			Washout Time (minutes)			Travel to Site Time (minutes)		
1	7	1	12.6	6	315	20.8	normal	10	1	lognormal	20	1	normal	5	1	beta	3	2	normal	5	1	lognormal	20	1
2	7	2	12.6	5	42	19.1	lognormal	12	2	lognormal	25	2	lognormal	7	2	lognormal	12	2	lognormal	7	2	lognormal	25	2
3	7	2	12.6	3	48	19.1	gamma	3	2	gamma	7	6	gamma	3	2	gamma	5	7	gamma	3	2	gamma	7	6
4	7	2	12.6	3	46	14.8	weibull	4	2	weibull	9	5	weibull	4	2	weibull	4	2	weibull	4	2	weibull	9	5
5	7	2	12.6	3	66	13.1	uniform	5	10	uniform	20	30	uniform	5	10	uniform	5	10	uniform	5	10	uniform	20	30
6	8	2	5.5	5	48	15.5	beta	5	1	beta	12	3	beta	5	1	beta	5	1	beta	5	1	beta	12	3
7	9	1	5.5	6	231	16	exponential	2	6	exponential	5	10	exponential	2	6	exponential	2	6	exponential	2	6	exponential	5	10
8	9	1	5.5	7	227	18.6	normal	10	4	normal	22	8	normal	9	4	normal	9	4	normal	9	4	normal	22	8
9	9	2	10.9	4	50	13.8	beta	5	2	beta	7	5	beta	5	2	beta	5	2	beta	5	2	beta	7	5
10	10	1	10.9	5	51	14.2	lognormal	8	5	normal	18	7	lognormal	6	5	lognormal	6	5	lognormal	6	5	normal	18	7

Figure 8.3: A sample of the case base from *CBRSim*

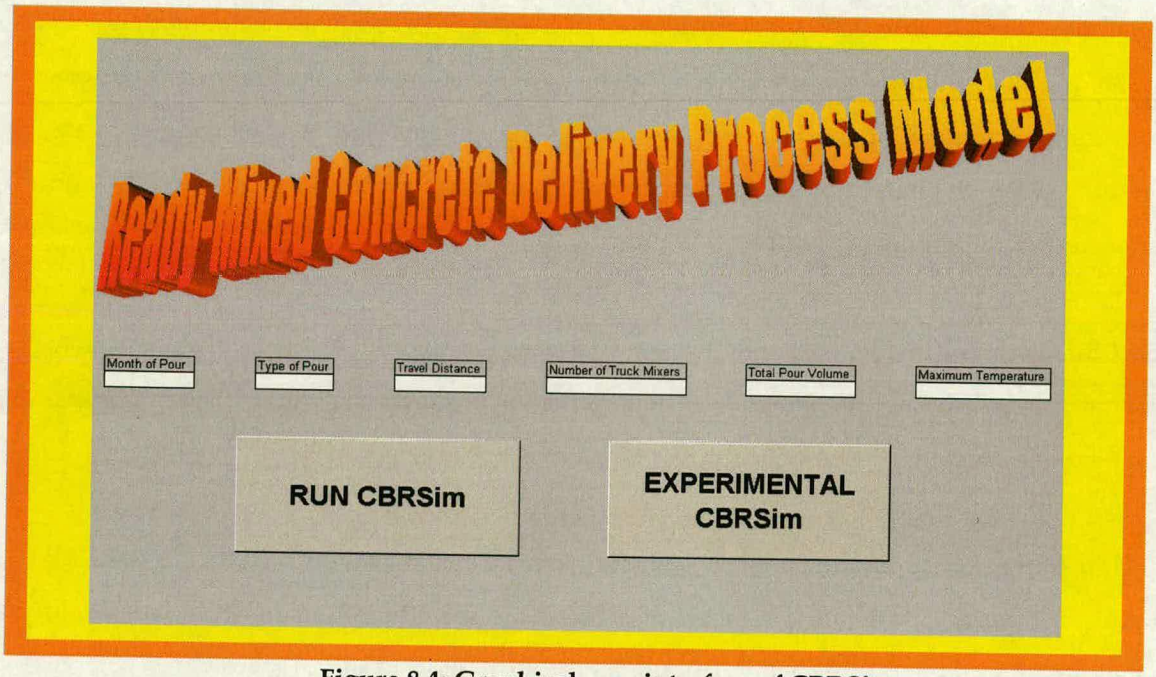


Figure 8.4: Graphical user interface of CBRSim

8.2.2 *DES module*

The DES module (when not performing experiments – see chapter 9) has a number of assumptions built in. They are:

- There is only one concrete pump present on the project site.
- There is one ganger, one pump operator, two general labourers and two concrete finishers involved in the operation, on the project site.
- The cost of the RMC material is not incorporated into the value direct cost (see appendix E for the full calculation of this value).

Once the CBR module has produced an output, its role is at an end (see figure 8.1). This output, in the form of the probability distributions (and parameters) that represent the 6 activities within the RMC supply process, then becomes part of the input to the discrete-event simulation module. The remainder of this input comes from the user-defined information first

presented to the CBR module (see figure 8.1) and includes information that is important to the DES module. For example, the total pour volume tells the DES module when to stop simulating.

The DES module works in the same manner as the discrete-event simulation model from chapter 6. To reiterate this involves:

- Generating 1000 estimates of the duration of each activity (random variates) within the RMC supply process from the probability distributions suggested by the CBR module.
- Averaging these 1000 random variates to produce an average duration for each of the activities mentioned above.
- Calculating the time that the next event should occur. Where an event is either an arrival or departure of a truck mixer from the batching plant, concrete pump or the washout area.
- Recording the amount of RMC that has been delivered on each occasion of a truck mixer arrival event at the concrete pump.
- When a truck mixer completes a cycle (i.e. gone from batching plant back to the batching plant via the project site) steps 1 to 4 should be repeated, until the amount of RMC delivered in the simulation equals the amount stipulated by the *CBRSim* user.
- When the simulation comes to an end. A summary of the key statistics, such as productivity, direct cost, duration and truck mixer idle time, etc., is presented in a tabular form for the perusal of the user. Note the mathematical formulae used to calculate these statistics is shown in appendix E.

8.3 Model Validation

The process of validation involves ensuring that any model developed is an accurate representation of the system or process that is being modelled. Here, the method of validation used in chapters 6 and 7 was used again. To recap, this method is to compare the model output (process performance measures) with data from the observed process.

In section 8.3.1 the *CBRSim* model is validated. Subsequently, a comparison is made between *CBRSim*, the discrete-event simulation model (of chapters 6 and 7) and the CBR model 10 (of chapter 7). Finally, some limitations of *CBRSim* are discussed.

8.3.1 CBRSim model validation

A data set of 50 concreting operations (from 4.5) that was not used in the development of any models was retained for use in the validation process. Each of these observed operations were recreated by the *CBRSim* model, resulting in estimates of the process performance measures, given the observed operating conditions (e.g. total pour volume, etc.). These estimates were compared with the observed process performance measures, over the entire validation data set. A graphical display of this comparison is shown for productivity in figure 8.5 and for duration in figure 8.6. The 1:1 line is indicative of a model that has provided accurate estimates of reality (observed values). Clearly, *CBRSim* was found to be capable of producing estimates of these two process performance measures with a high degree of accuracy.

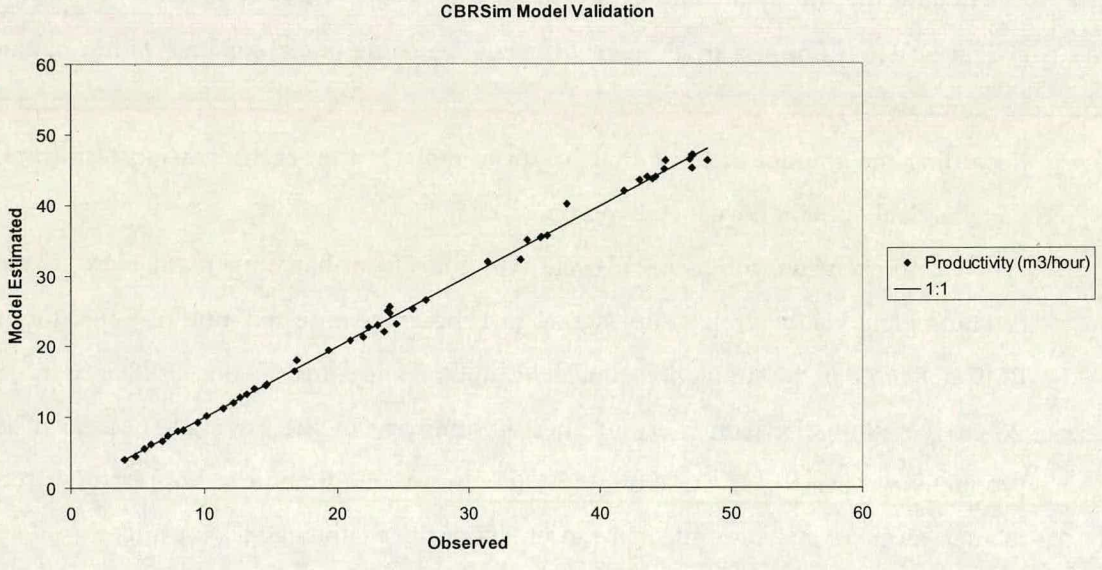


Figure 8.5: *CBRSim* estimated vs. observed productivity

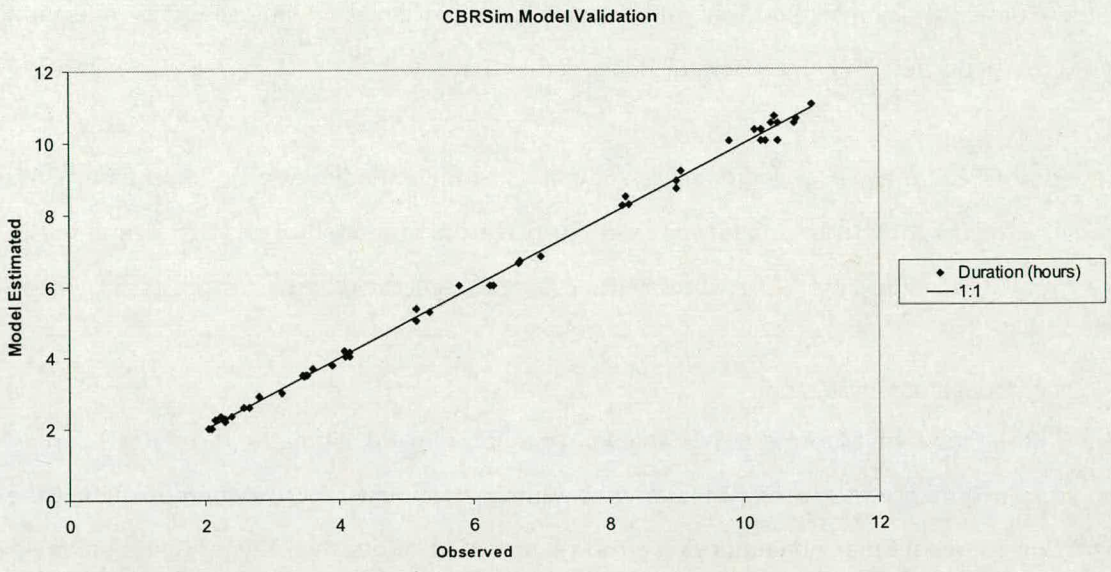


Figure 8.6: CBRSim estimated vs. observed duration

Table 8.1 displays the summary statistics of the validation process of *CBRSim*, where each value is in terms of the percentage difference between the model estimate and an observed value.

Statistic	Process Performance Measure				
	Productivity	Duration	Direct Cost	Truck mixer Utilisation	On-Site Resources Utilisation
Max	7	5	8	14	17
Min	-7	-5	-9	-12	-18
Mean	2	1	-1	1	-1
75th Percentile	1	3	4	8	8
25th Percentile	-2	-1	-5	-6	-9

Table 8.1: Summary statistics of *CBRSim* validation (%age difference between model estimated and observed process performance measure)

CBRSim was found to be capable of producing accurate estimates of direct cost, truck mixer utilisation and the utilisation of on-site resources with +/-20% accuracy, at its worst. The low mean value of the difference between estimated and observed has significance, as the average result from the model is extremely accurate. This fact indicates that CBRSim on the main produces accurate estimates of these process performance measures, but encountered a small number of validation observations that CBRSim found difficult to estimate. It is the estimates

of these observations that produced a fluctuation in, for instance, on-site resources utilisation of +17 to –18 percent difference from the observed.

Now that CBRSim has been validated, a comparison can be made with the Excel-based DES model, *ExcelSim* (of chapter 6) and the case-based reasoning model (of chapter 7) in a bid to determine if the hybrid model provides better estimates than the others.

8.3.2 Model cross-validation

The *CBRSim* model was found to be capable of providing valid estimates of the RMC supply process performance measures. This notwithstanding, it is important to examine whether the extra development time and computing effort required in *CBRSim* over the individual models: the discrete-event simulation (*ExcelSim*) model of chapters 6 and 7; the case-based reasoning (CBR) model 10 of chapter 7. If either of these models is capable of providing estimates as accurately as *CBRSim*, then the hybrid model is clearly not the solution to the complexity problem briefly mentioned as the beginning of this chapter. Rather, an alternative method of reducing the statistical knowledge required by a construction practitioner in construction process modelling will need to be found.

The *ExcelSim* and CBR models were validated using the same validation data set discussed in 8.3.1. Thus the results between the *CBRSim*, *ExcelSim* and CBR models are comparable. Figures 8.7 (productivity), 8.8 (duration), 8.9 (direct cost), 8.10 (truck mixer utilisation) and 8.11 (on-site resources utilisation) show the validation results, which are the (average and maximum) errors produced by the three models in estimating the process performance measures for the entire validation data set. The smaller the average error the more accurate the model. Also the smaller the maximum error the more reliable the model. Using these generalisations the results shown in figures 8.7 to 8.11 have been analysed.

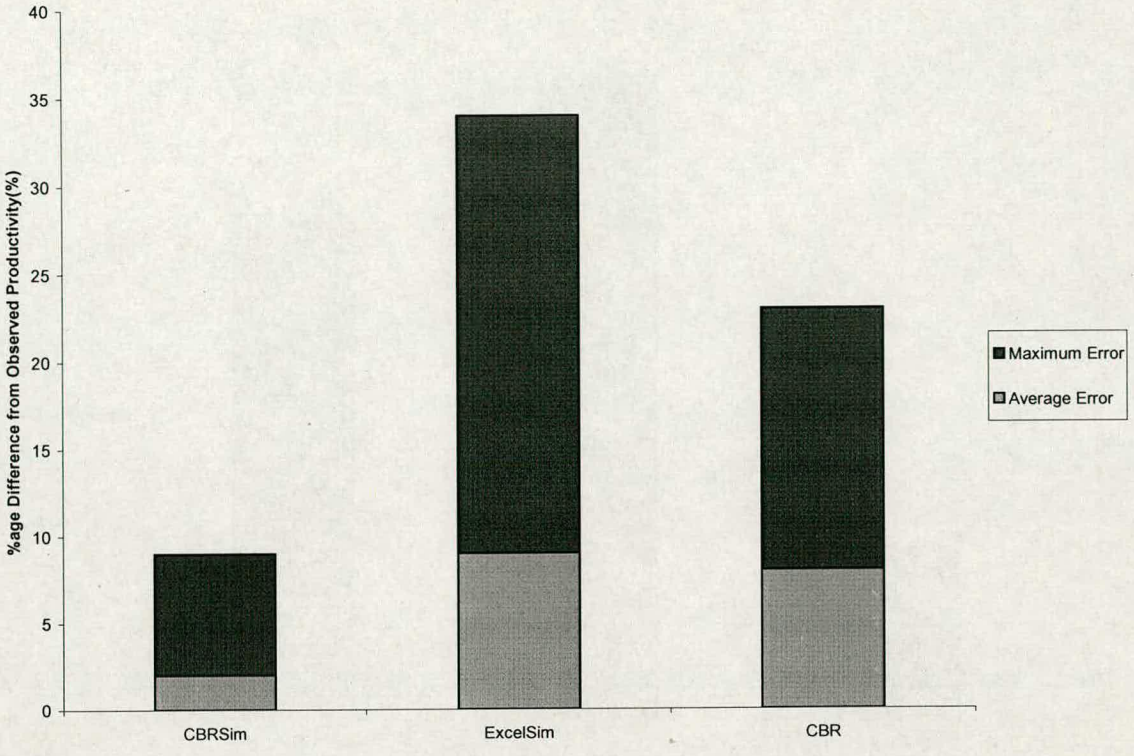


Figure 8.7: Errors produced by each model in estimating productivity

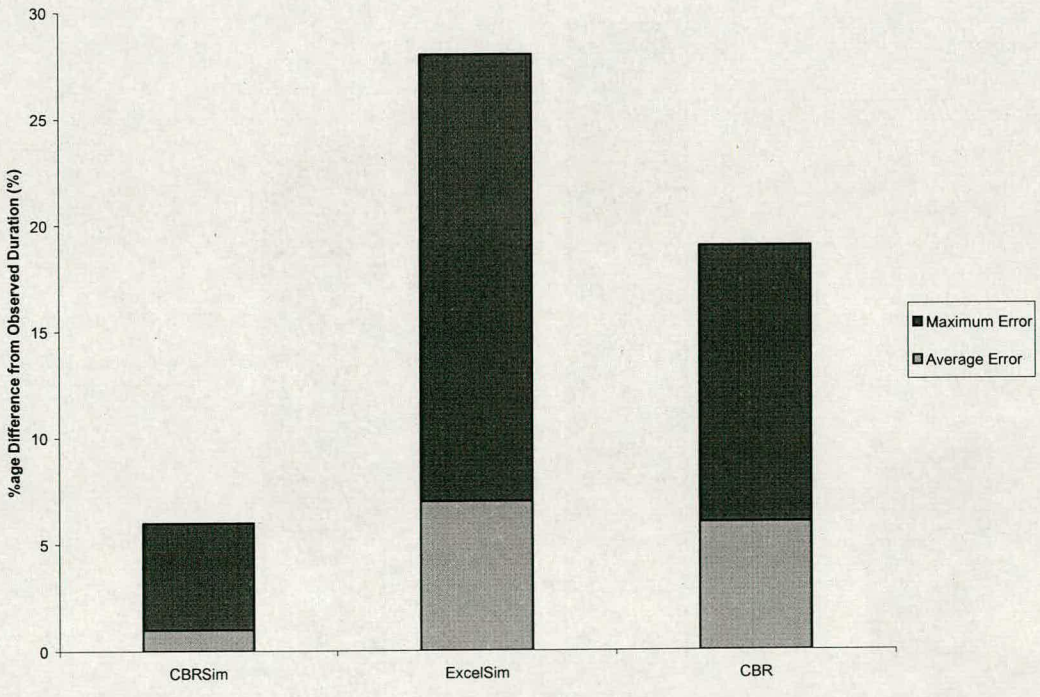


Figure 8.8: Errors produced by each model in estimating duration

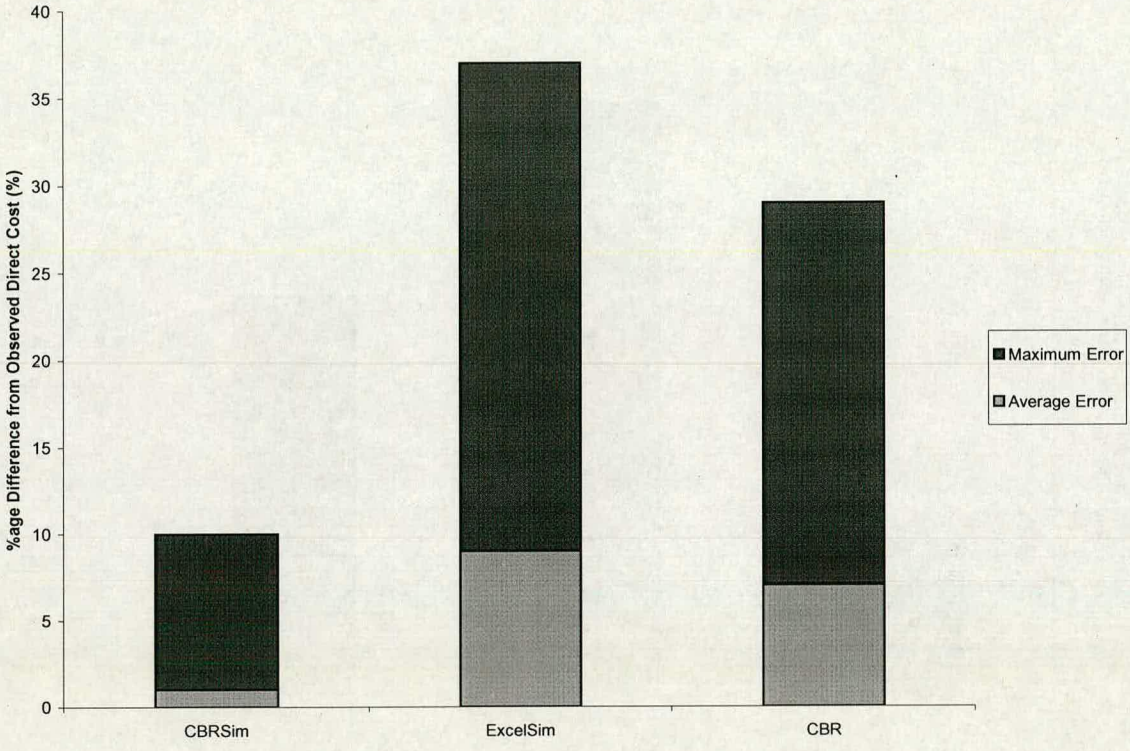


Figure 8.9: Errors produced by each model in estimating direct costs

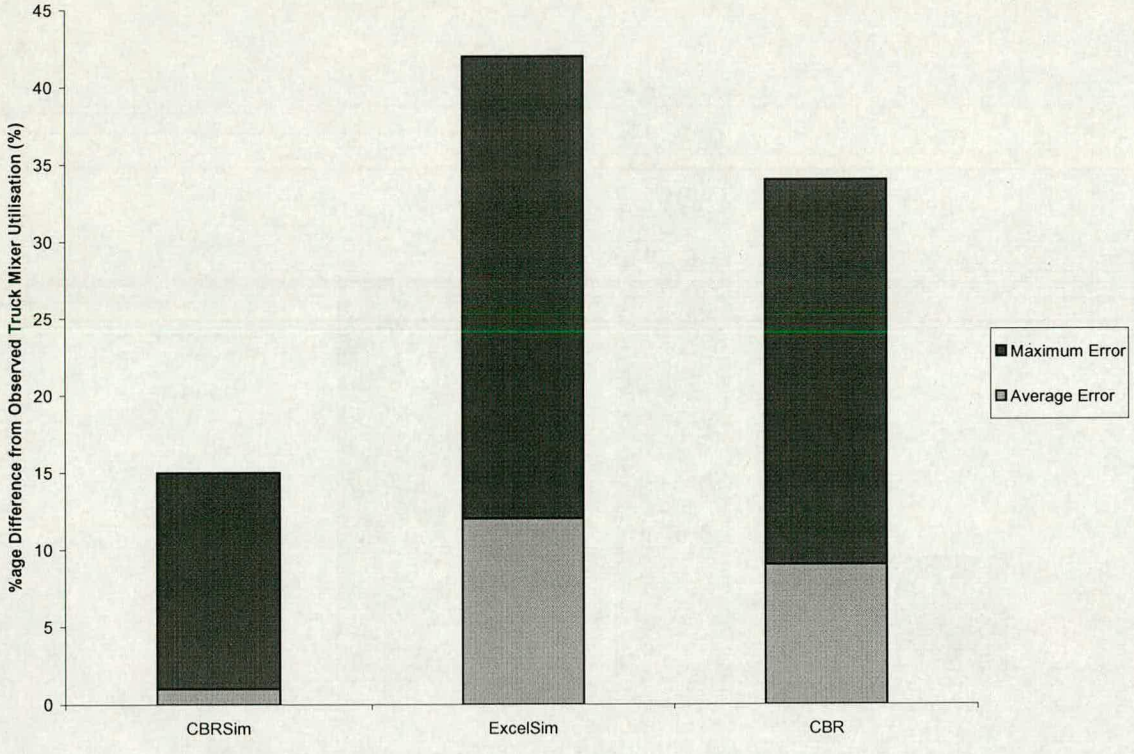


Figure 8.10: Errors produced by each model in estimating truck mixer utilisation

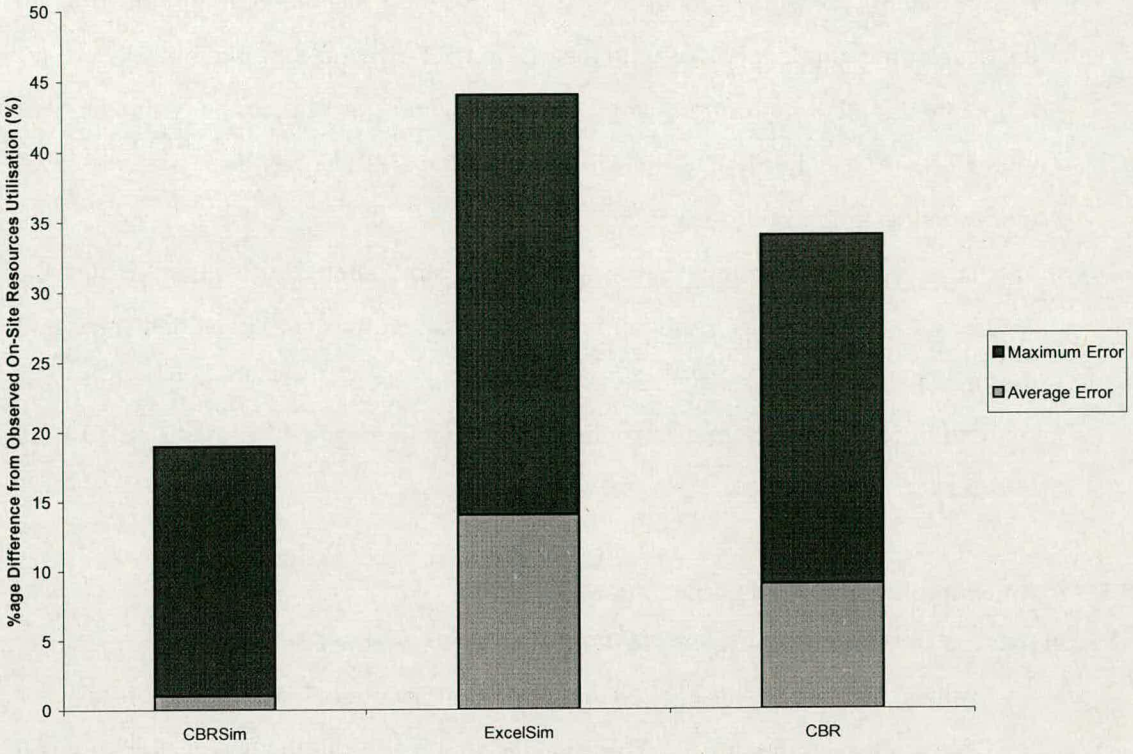


Figure 8.11: Errors produced by each model in estimating on-site resources utilisation

Figures 8.7 to 8.11 show that for all of the process performance measures, *CBRSim* provided more accurate estimates (lower mean error) with more reliability (smaller maximum error) than either the *ExcelSim* or CBR model. Thus there is no need to seek an alternative method of modelling the RMC supply process. Although *CBRSim* provided accurate estimates of the RMC supply process performance measures most of the time, it was not capable of achieving 100%, with 100% reliability. Thus, a construction practitioner must be aware of the limitations of *CBRSim*. These limitations are discussed next.

8.4 Limitations of *CBRSim*

Any model should be treated with a certain level of suspicion, regardless of the accuracy produced in a validation study. It is for this reason that a list of limitations of the *CBRSim* model has been drawn up. This list is by no means exhaustive, but could include:

- Each run of a stochastic simulation model produces only estimates of a model's true characteristics of a particular set of input parameters. This problem has been mitigated to a certain degree because *CBRSim* utilises 1000 estimates in to produce an output.

- Simulation models are often expensive and time consuming to develop (Law and Kelton, 2000). However, it is hoped that by presenting this thesis in a public domain the development time involved with regards to *CBRSim* would be minimised.
- It may take some time commitment for a construction practitioner to become familiar with *CBRSim*. It is hoped that the simple way that *CBRSim* has been presented will minimise this time commitment.
- The large volume of numbers produced by a simulation study often creates the tendency to place greater confidence in a study's results than is justified (Law and Kelton, 2000). If a model is not a "valid" representation of a process under study, the simulation results, no matter how impressive they appear will provide little useful information about the actual process.

8.5 An example to illustrate how to use *CBRSim*

The following is an example which aims to illustrate how to use the *CBRSim* model:

- The values that are to be entered into the *CBRSim* model via the graphical user interface are shown in table 8.2. These values are taken from the observation of a real concreting operation.

Variable	Value
Month of Pour	March
Type of Pour	Wall
Travel Distance (miles)	18
Number of Truck mixers	4
Total Pour Volume (cum)	80
Maximum Temperature (oC)	12

Table 8.2: Input values of the variables of the CBRSim model

- Having entered the required information, the user must now click on the *RUN CBRSim* button (see figure 8.4). Clicking this button results in both the CBR (as described above) and the DES (also described above) processes being undertaken.

Although the user does not have an opportunity to view the probability distributions (and parameters) suggested by the CBR module it is possible to retrieve this information. In the example concreting operation case 178 was suggested by CBR as a solution to the situation shown in table 8.2. The probability distributions and parameters held in case 178 are shown in table 8.3.

- The discrete-event simulation model stopped when the volume of RMC delivered to the concrete pump reached 80 m^3. The total time recorded for *CBRSim* to produce a result was 180 seconds. The estimates of the process performance measures produced by *CBRSim* are shown in table 8.4.
- These estimates could be used to produce a 'look ahead' construction plan.
- Discrete-event simulation is a flexible form of modelling and *CBRSim* is capable of performing experiments that can provide a construction practitioner with much more knowledge of the RMC supply process under certain conditions than a mere estimate of the performance process measures. This experimental capability is explored in chapter 9, where experiments are outlined that aid a construction practitioner to make management decisions about the process.

Case Number	Activity Duration											
	Batch Time		Travel to Site Time		Position Time		Discharge Time		Washout Time		Travel from Site Time	
178	lognormal	1.5	normal	20 2	lognormal	8 1	gamma	3 2	lognormal	4 2	normal	20 3

Table 8.3: The distributions (and parameters) suggested by CBR to solve the example

Process Performance Measures				
Productivity (m3/hour)	Duration (hours)	Direct Cost (Pounds)	Utilisation of Truck Mixers (%)	Utilisation of On-Site Resources (%)
12	6.67	1200	95	63

Table 8.4: Process performance estimates produced by CBRSim in the example

8.6 Chapter conclusions

In this chapter a hybrid model, *CBRSim* capable of producing estimates of the RMC supply process is presented. It was found that this model could produce accurate estimates, but encountered a few cases that were difficult to estimate.

A cross validation between *CBRSim*, a simulation model (*ExcelSim*) and a CBR model (model 10) was performed to determine if the extra development and operational effort attached to *CBRSim* was worthwhile. It was found that *CBRSim* had a smaller average error in validation than the other two models, inferring that it is more accurate at estimating the RMC supply process. Also, *CBRSim* had a much smaller maximum validation error, showing that it is more reliable than the other two models.

CHAPTER 9

CBRSim AS A DECISION SUPPORT TOOL

This thesis has so far concentrated on the detailed development of a robust, accurate and effective model of the RMC process. But this thesis must also investigate how this model can be used in practice and fulfil the three aims set down in chapter 2.

9.1 Introduction

In chapter 8 a model, *CBRSim* was introduced which removes the need for a construction practitioner to possess statistical knowledge of construction processes; thus allowing a practitioner to take advantage of the ability of simulation to provide accurate estimates of such processes. This chapter aims to show that simulation and hence *CBRSim* can be used to provide information through a series of quickly performed experiments that may aid decision-making. Initially the use of *CBRSim* to aid one of the most basic decisions which is (usually) made by a construction practitioner when ordering RMC is examined; this decision is how many truck mixers should be used to deliver the RMC. Then the capability of *CBRSim* to aid a more complex decision – that of who should be the supplier of RMC to a project, given a number of alternatives – is examined. Finally, some conclusions are drawn from this chapter.

9.2 Decision 1: How many truck mixers are required?

One of the key pieces of information, which a construction practitioner is usually required to provide (within a specified lead time) to a RMC supplier is the number of truck mixers that are required to perform a particular operation. By the traditional method of intuitively estimating this number it is difficult to fully understand the implications of requesting a particular number of truck mixers.

Discrete-event simulation is a flexible tool, which permits changes in the values of variables (truck mixers in this case) to be used in a simulation model and the results of these changes on the process performance measures. *CBRSim* shall be used to provide such information thereby simplifying this decision. Subsequently a brief analysis of this information shall be performed to identify the trends displayed in the results from the *CBRSim* experiment. Finally, the decision of how many truck mixers should be used in the particular operation is made taking into account the current pressures being placed upon the project (i.e. time is of the essence).

The experiment required to aid this particular decision is to be known as 'the truck mixer experiment' and it involves:

- Entering the information of an example case (contained in table 9.1) into the *CBRSim* model, with the addition of the range of truck mixers. For the first run the number of truck mixers to be used will be 3, increasing in increments of 1 until there are 8 truck mixers being used in an operation.

- The *CBRSim* model then performs 6 full simulation runs (one for each number of truck mixers) and consolidates the results into a table. This information is not presented to the user in tabular form, rather it is converted into a series of graphs that make the information simple to analyse.

Variable	Value
Month of Pour	March
Type of Pour	Wall
Travel Distance (miles)	18
Total Pour Volume (cum)	80
Maximum Temperature (oC)	12

Table 9.1: Input values of the variables of the *CBRSim* model

9.2.1 Results from the truck mixer experiment

Although experimental results are not presented in *CBRSim* in tabular form, they have been included here for completeness (in table 9.2). When presented in this form it is difficult to determine the number of truck mixers that produce a good productivity value and a low cost. When the same results are presented in graphical form such a task is simplified. For this reason the results displayed in table 9.2 are shown in graphical form in figures 9.1a - 9.1d.

Input	Process Performance Measures				
Number of Truck Mixers	Productivity (m3/hour)	Duration (hours)	Direct Cost (Pounds)	Utilisation of Truck Mixers (%)	Utilisation of On-Site Resources (%)
3	7	11.4	1550	100	56
4	12	6.7	1200	95	63
5	14	5.7	1100	89	68
6	16	5.2	1020	81	78
7	18	4.4	1200	69	88
8	20	4.1	1480	65	94

Table 9.2: Results of the truck mixer experiment

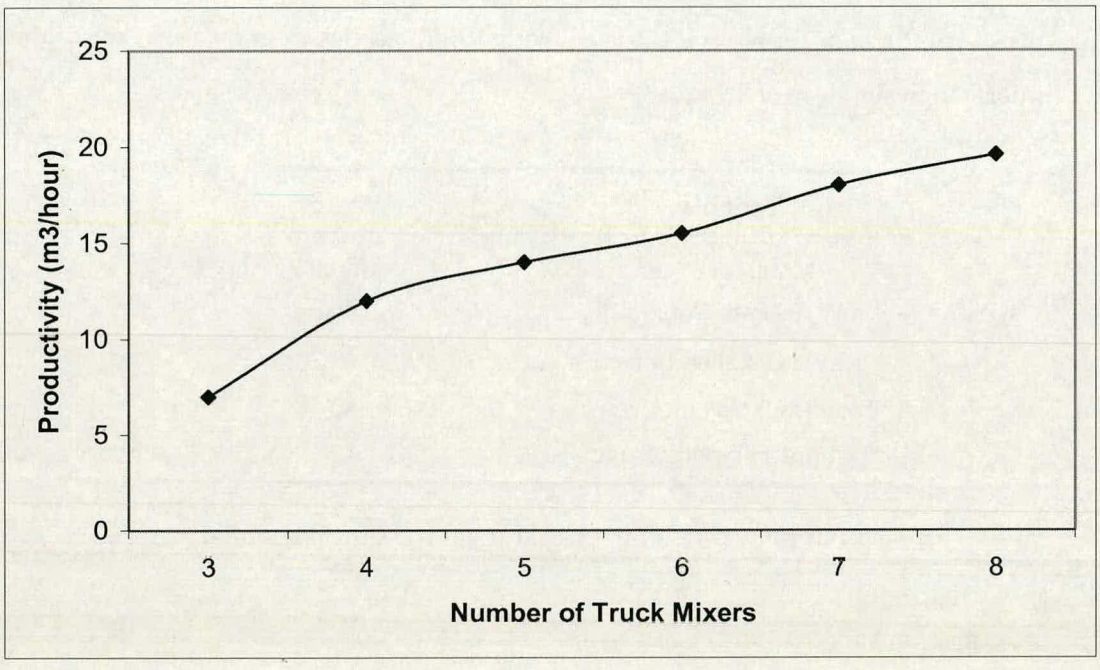


Figure 9.1a: Productivity estimates from the truck mixer experiment

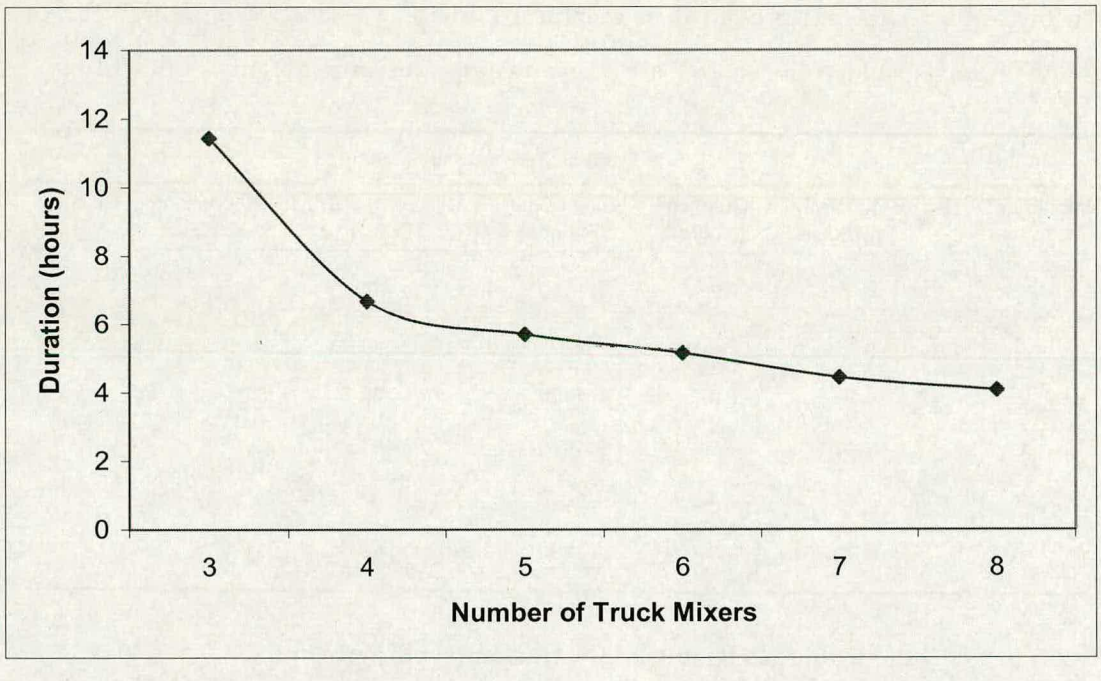


Figure 9.1b: Duration estimates from the truck mixer experiment

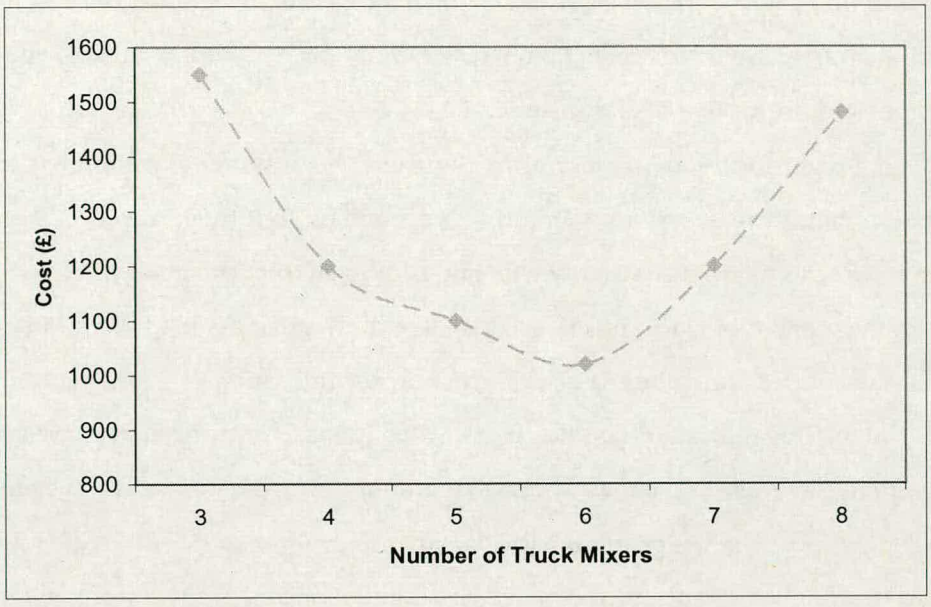


Figure 9.1c: Cost estimates from the truck mixer experiment

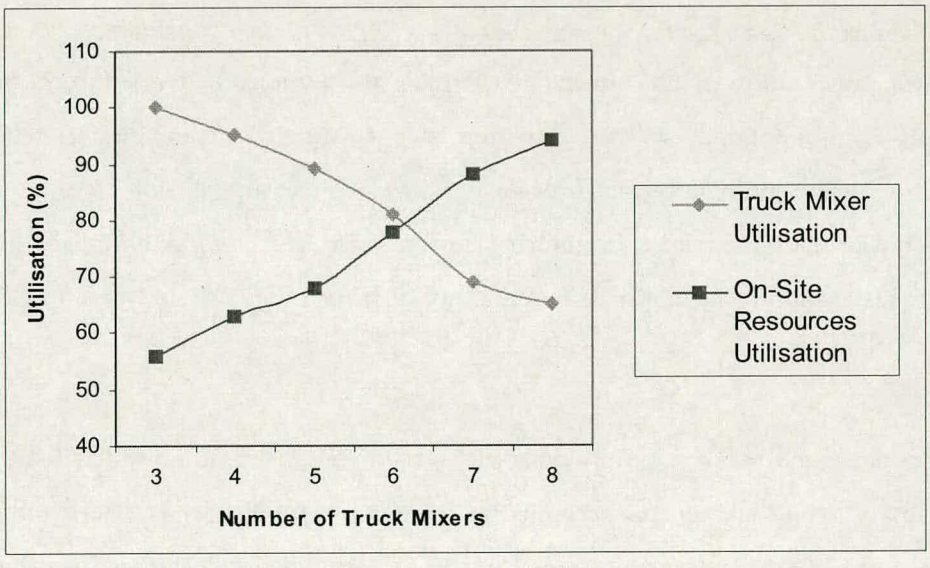


Figure 9.1d: Resource utilisation estimates from the truck mixer experiment

9.2.2 *A brief analysis of the experimental results*

It may be useful to identify the trends contained within the results produced by the *CBRSim* experiment and this can be done by performing a brief visual analysis of figures 9.1a – 9.1d. The findings of such an analysis can be summarised as:

- Productivity rises as the number of truck mixers used in an operation increases.

- The duration of an operation decreases with an increase in the number of truck mixers used in an operation. This reduction occurs rapidly for 3 to 5 truck mixers, and then declines steadily for 5 to 8 truck mixers.
- Costs are high (for a low number of truck mixers) because the operation has a high duration, but decreases to a minimum (corresponding to 6 truck mixers). The trend then is for costs to increase steadily with any additional truck mixers.
- When the number of truck mixers used is 3 there is a great demand placed upon the truck mixers and thus there is a high truck mixer utilisation of 100% (i.e. they are never idle). This utilisation rate decreases as the number of truck mixers used in the operation is increased. The reverse may be said for the on-site resources – when there are 3 truck mixers in the operation, the on-site resources are under-utilised. However, when the number of truck mixers is increased in the operation there is more demand placed upon the on-site resources, resulting in a higher utilisation rate.

9.2.3 Making the decision (on truck mixer quantity) given the project constraints

The graphical presentation of the impact of changing the number of truck mixers on the process performance measures, has provided the construction practitioner with an opportunity to easily optimise certain process performance measure(s). For example, if the duration of the operation recreated in the truck mixer experiment is to be optimised (in this case minimum duration being optimal) the quantity of truck mixers should be 8 (see figure 9.1b).

The decision on which process performance measure(s) should be optimised will be most effective if made while taking into account the current constraints being placed upon the project. For example a project's main constraint may be cost as a result of being over-budget, and the cost of RMC operations should be optimised; or a project may be late, resulting in a need for fast operations, hence time is the major constraint in this instance, and the duration of RMC operations should be optimised.

Armed with the knowledge of what the major constraint of a project is, a construction practitioner could use the graphical output of the truck mixer experiment to choose the quantity of truck mixers that helps to reduce the effect of the constraint on the entire project. For instance, if cost is a major constraint to a project then RMC operations performed in a

manner that minimises cost will help reduce the effect that cost has on the whole project. By planning RMC operations in this manner, the practitioner is looking beyond the obvious and immediate requirement that an individual operation be optimised, and is planning for the better good of the whole project (this is a concept of lean construction).

Next a number of common project constraints are drawn up and the likely responses to maximise the process performance given these constraints are presented.

Constraint One: A project is late

If a project is late, time is obviously of the essence. Thus the duration of RMC operations should be minimised and either the productivity (m^3/hour) or duration (hours) graphs are of primary interest in this case. Here, the duration graph (figure 9.1b) shall be examined and it is presented again in figure 9.2a. Clearly, using 8 truck mixers in the operation provides a minimum duration of 4.1 hours.

Having the information presented in this manner highlights the potential impact of not maximising the process performance given the current project constraints. It would not be unrealistic that 3 truck mixers may be utilised in this instance (4 truck mixers were really used in this case) and this would result in an operation of 11.4 hours duration. For a project under tight time constraints these additional 7.3 hours (11.4 – 4.1) over one operation are important, but if this trend were to be witnessed over a number of operations the time spent performing RMC operations could become crucial to a project's success.

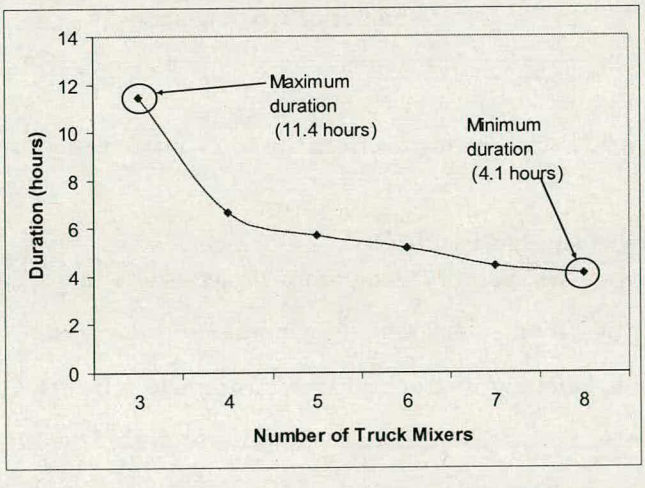


Figure 9.2a: Duration estimates from the truck mixer experiment

Constraint Two: A project is over budget

In a project where cost is the most prominent factor, then a focus should be placed upon the cost graph (figure 9.1c). This graph is presented again in figure 9.2b and from this it can be seen that a minimum cost for the operation of £1,020 is achieved by using 6 truck mixers.

To compare the extreme outcomes of performing this operation, if 3 truck mixers were used then the operation would cost £1550. The increase of £530 caused by using 3 truck mixers instead of 6 is a large difference for what is a small RMC operation, and the consequences of such an error could be much more significant if made over a number of operations or in a larger RMC operation.

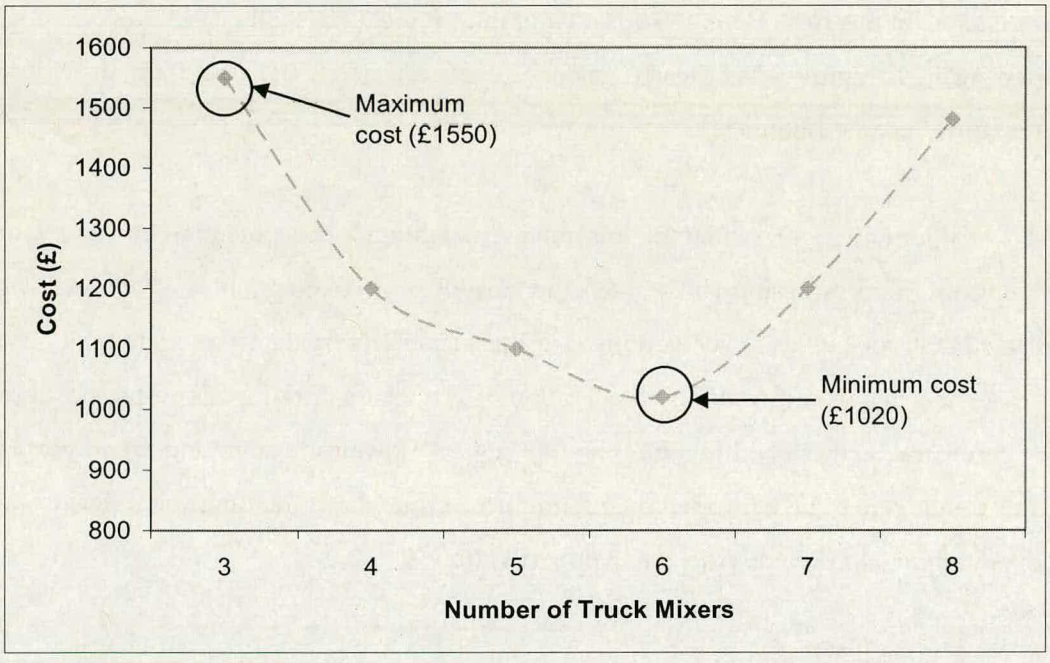


Figure 9.2b: Cost estimates from the truck mixer experiment

Constraint Three: Project site space is limited

It is entirely feasible that the space available on a project site is a governing factor for the project (e.g. city centre developments) and in this situation it is implausible to expect that there could be much (or any) queuing of truck mixers on the site. Thus the resource utilisation graph (figure 9.1d) is of importance in this instance and is presented again here as figure 9.2c.

In order to minimise the space required for the performance of the RMC operation 3 truck mixers should be used. This would result in a truck mixer utilisation rate of 100%, which translates to no queuing of truck mixers on a project site.

Clearly these situations have been simplified, as there is rarely a situation where only one factor is of primary importance, but these examples illustrate the potential of *CBRSim* as a practical tool to aid simple decision-making in the planning of RMC operations.

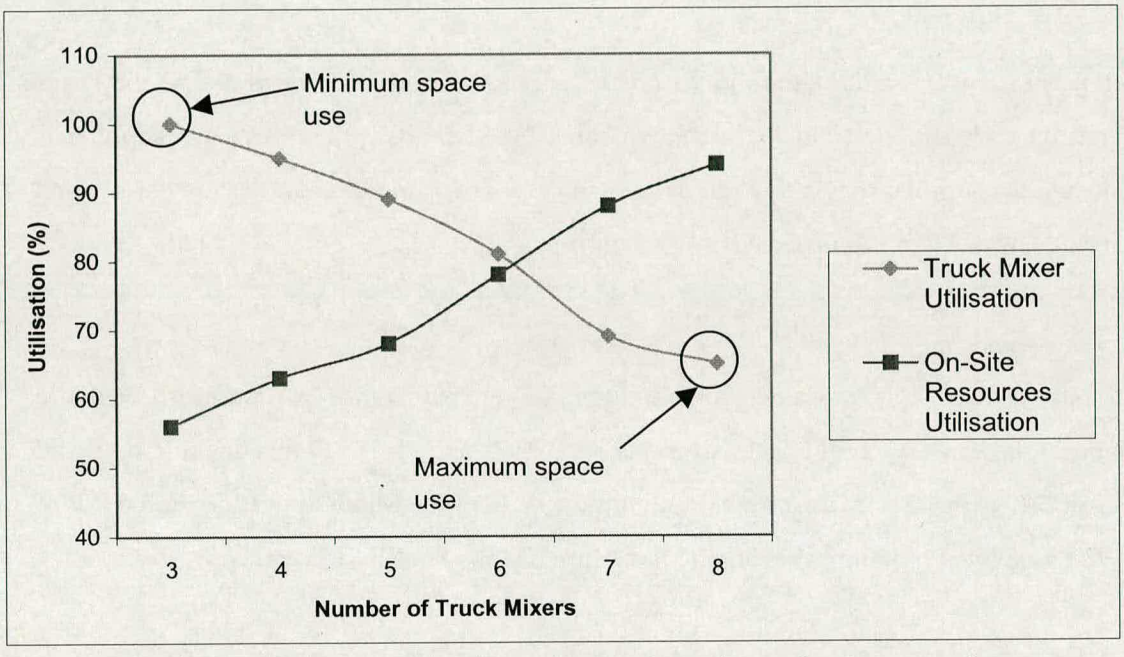


Figure 9.2c: Resource utilisation estimates from the truck mixer experiment

9.3 Decision 2: RMC supplier selection

In this section *CBRSim* shall be used to aide the decision regarding which company should be appointed as the supplier of RMC to a project.

A RMC supplier's tender bid for a project usually only includes a unit cost for the RMC (£/m^3) and a likely delivery rate (m^3/hour). From the perspective of a contractor this is not the only cost that will be incurred in RMC operations – there are costs associated with having a concrete pump on the project site and workers to help place the RMC – the on-site resources. These costs are a function of the duration of a RMC operation and should be added on to the basic bid provided by a RMC supplier, to provide a more realistic assessment of cost. *CBRSim* is capable of producing this realistic cost estimate and by taking into account the stochastic

nature of the process can provide a more realistic estimate of the delivery rate of RMC. These capabilities are used to provide accurate information that will permit a fair decision on which company should be selected as the RMC supplier.

To begin the project details shall be stipulated, followed by a description of the terms that the RMC suppliers in the area were offering. The project is an American one and hence the material volumes are measured in cubic yards and costs are measured in dollars (m^3 and £[2] equivalent provided in brackets).

CBRSim is used to simulate the impact on the process performance measures of the terms offered by each supplier. Finally, a comparison of the services offered by each supplier is made and the supplier that is thought to be capable of providing the best service, based upon the results from *CBRSim*, is selected.

9.3.1 Project details

The construction of the Transamerica Pyramid building in San Francisco, California (see figure 9.3) began in 1969. The building was completed in 1972 and stands to this day at a height of 853 feet (240 m) which made it the tallest building west of the Mississippi River in the U.S.A. in 1972; today it is 4th tallest west of the Mississippi (i.e. west of Chicago and New York).

This case study is going to focus on the cost of construction of the base floors of the pyramid building, because they were formed using RMC. The base floors comprise a car park (280 spaces), retail outlets and restaurants. The volume of RMC required to construct these floors totalled 16,000 yd^3 (12,200 m^3)(The Pyramid Center, 2005) and the operations required to place this RMC were 34 x 300 yd^3 (230 m^3) base pours, 40 x 100yd^3 (75 m^3) wall pours and 36 x 50yd^3 (40m^3) column pours – this information was supplied by Mr. Bill Garland, then concrete manager of Dinwiddie Construction Co. (now manager of Bode Cement, San Francisco, CA).

A quote was acquired for the pours mentioned above from three local RMC suppliers. Although this building was constructed around 1970, the information from the RMC suppliers (e.g. costs) is based on 2005 prices. For confidentiality reasons the RMC suppliers (2 of the 3)

[2] $ to £ conversion made using exchange rate of 1.81 dollars to the pound sterling on 10/05/2005

wished to remain anonymous and thus all the suppliers are simply referred to as companies A, B and C in this thesis, to respect their wishes.

Figure 9.3: The Transamerica Pyramid, San Francisco, California, U.S.A.

9.3.2 Supplier details

The San Francisco area contains a number of RMC suppliers and the competition to win contracts for these companies is fierce. For this project, three RMC suppliers of upstanding reputation were identified as being within a 30-minute driving distance of the project site and they were asked to submit a bid for the project. The details of the tender bids submitted by each of these companies are shown in table 9.3.

Company	Distance from Project (miles)	RMC Unit Cost in $/yd3 (£/m3)	No. truck mixers			Delivery Rate in yd3/hour (m3/hour)		
			Base	Wall	Column	Base	Wall	Column
A	19	54 (39)	6	3	3	27 (21)	15 (11)	15 (11)
B	16	54 (39)	Values requested from customer			Not specified		
C	24	56 (40)	7	4	4	28 (21)	16 (12)	16 (12)

Table 9.3: Tender bids submitted by the RMC suppliers (£ and m³ in brackets)

Company C have offered a higher rate of delivery for each pour type and more truck mixers than offered by Company A, but at a higher unit cost of RMC. Company B have requested that the customer place an order for the number of truck mixers required for each pour type. Incidentally, Company B's unit cost matches the low bid provided by Company A.

From table 9.3 it is difficult to determine which bid may end up costing the contractor more money and time. In the next section *CBRSim* will be used to simulate the impacts of each of these bids on the process performance measures, thus providing a more realistic and comparable measure of cost and duration than those provided in the tender bids.

9.3.3 *Simulating the realistic cost and delivery rate offered by each supplier*

As mentioned above there are more costs that must be met by a contractor in performing RMC operations than the basic unit cost of RMC. These additional costs are associated with the hiring of plant and workers and are therefore directly related to the duration of an RMC operation. Each of the 3 suppliers of RMC will produce a different total cost of RMC operations, because the durations of RMC operations will differ for each supplier. *CBRSim* is to be used in this instance to provide estimates of the likely total cost, duration, etc. produced by each supplier given the conditions they have placed upon this project (i.e. travel distance to the project).

The methodology used in this simulation study is as follows:

- For each supplier, the process performance (e.g. duration) that they are capable of delivering is simulated for each pour type. This is because the productivity that can be produced in a base pour is very different from that of a wall pour. Additionally, if one company were capable of producing a low cost/duration for only a specific type of pour, it may be possible to contract that company to undertake only that type of pour and select another to undertake the remainder.
- The method of simulating the process performance estimates is different in the case of Company B from the others. This is because Company B has not stipulated the number of truck mixers to be used in the RMC operations. This affords an opportunity to perform the 'truck mixer experiment' detailed in 9.2, to determine the impact on the process performance measures of varying the quantity of truck mixers. From this

impact assessment the quantity of truck mixers which affords the best performance can be easily chosen.

- In the cases of Companies A and C *CBRSim* is used to simulate the RMC operations under the terms offered in their respective contract bids.

9.3.4 *Simulation results*

Company B

To begin, the simulation results for the case of Company B shall be examined. As the 'truck mixer experiment' was undertaken for Company B the results shall be presented in a graphical form, to make a comparison of the different options (numbers of truck mixers) easy. Figures 9.4a – 9.4d display the results of this experiment, when performed for base pours.

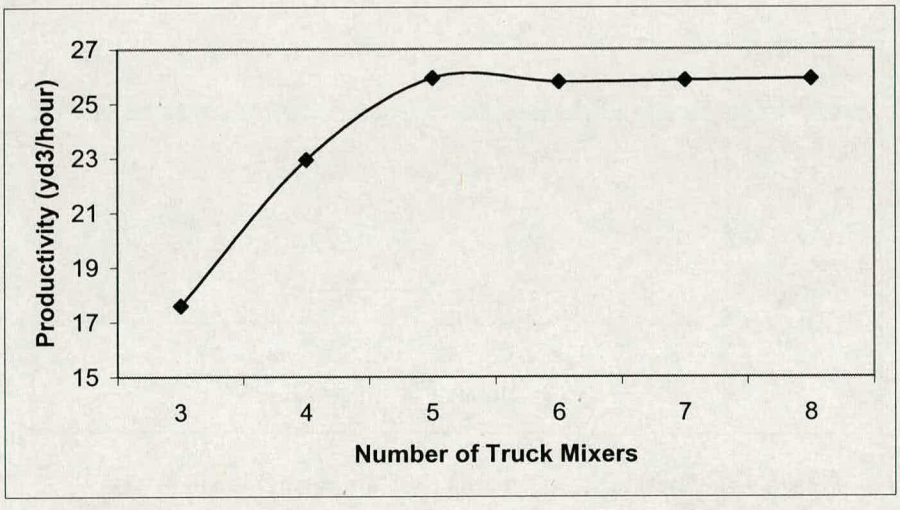


Figure 9.4a: Productivity estimates for Company B (base pours)

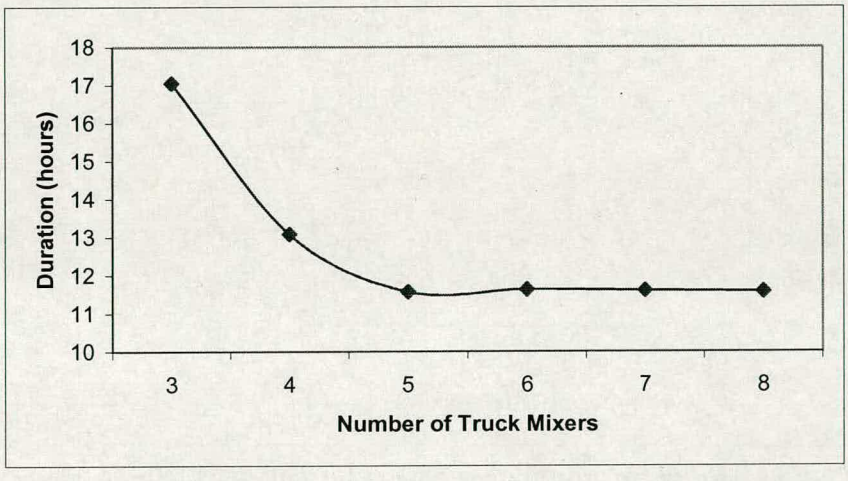


Figure 9.4b: Duration estimates for Company B (base pours)

The number of truck mixers that produces an optimum value for each process performance measure shall be assessed from each of the above graphs. This information will then be considered collectively to determine the quantity of truck mixers that will produce well performed (i.e. low cost, duration, etc.) base pour operations. For example in figure 9.4a estimates of how the productivity will vary with the number of truck mixers are shown. In this instance, 5 truck mixers produced an optimum productivity. This assessment was repeated for duration, direct cost and the resource utilisation graphs and the findings are summarised below in table 9.4.

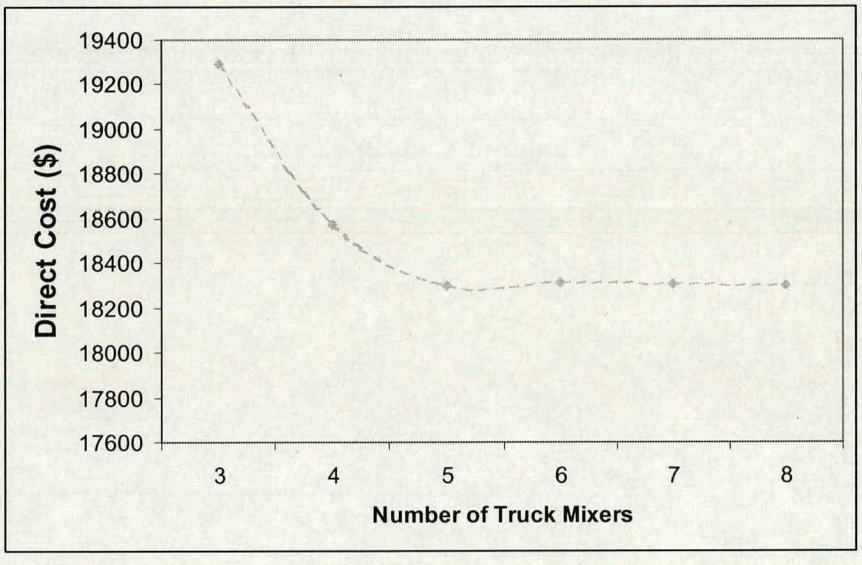


Figure 9.4c: Direct cost estimates for Company B (base pours)

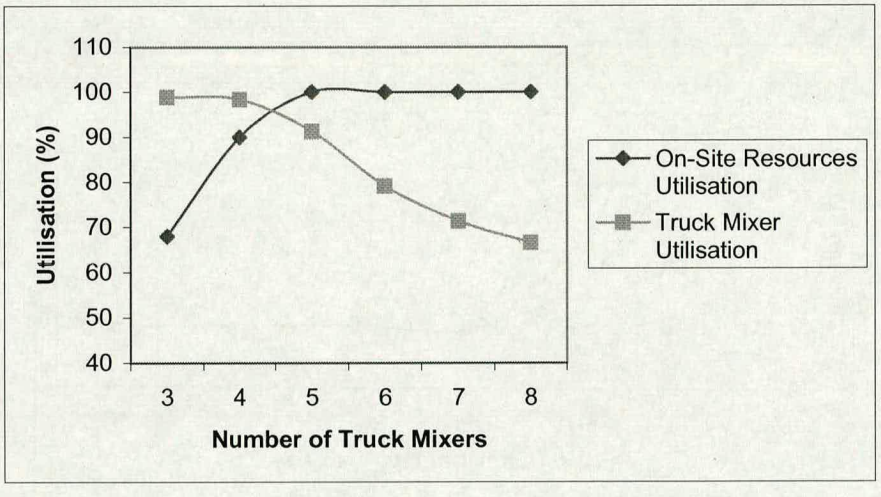


Figure 9.4d: Resource utilisation estimates for Company B (base pours)

Process Performance Measure	Project Constraint (as defined in 9.2)	No. of Truck Mixers that Optimise the Measure (Given the Constraint)
Duration	1	5
Direct Cost	2	5
Truck Mixer Utilisation	3	3

Table 9.4: Assessment of estimates for Company B (base pours)

Clearly, if time (constraint 1 defined in 9.2) and cost (constraint 2) are the governing constraints to this project then 5 truck mixers should be used in any base pour operations performed by Company B. If the project site is extremely limited for space (constraint 3), the truck mixer utilisation should be considered as the governing project constraint. However, this building's base is spread over a large area and space is not a governing constraint, thus it should be communicated to Company B that 5 truck mixers should be used in their base pour operations.

This procedure was repeated for typical wall and column RMC operations and it was found that 4 truck mixers should be used in both these types of operations when undertaken by Company B. A summary of the process performance estimates (of productivity, etc.) produced by Company B for one operation of each type, given the use of the selected quantities of truck mixers is shown in table 9.5.

Pour Type	Volume in yd3 (m3)	Truck mixers	Productivity in yd3/hour (m3/hour)	Duration (hours)	Direct Cost in $ (£)	Resources Utilisation (%)	Truck Mixer Utilisation (%)
Base	300 (224)	5	26 (20)	12	18,300 (10,100)	100	91
Wall	100 (77)	4	22 (16)	5	6,200 (3,450)	97	98
Column	50 (37)	4	19 (14)	3	3,200 (1,800)	100	90

Table 9.5: Company B performance estimates

Companies A and C

The performance estimates produced by *CBRSim* for Company A and Company C are shown in table 9.6. The estimates produced in tables 9.5 and 9.6 are compared in the next section, firstly by examining the results for each process performance measure on an individual basis. Subsequently an assessment of the overall performance afforded by each supplier will be made and utilised to determine which supplier should be chosen in this project.

Company	Pour Type	Volume in yd3 (m3)	Truck mixers	Productivity in yd3/hour (m3/hour)	Duration (hours)	Direct Cost in $ (£)	On-Site Resources Utilisation (%)	Truck Mixer Utilisation (%)
A	Base	300 (224)	6	27 (21)	11	18,200 (10,000)	99	97
	Wall	100 (77)	3	14 (11)	7	6,700 (3,700)	63	99
	Column	50 (37)	3	14 (10)	4	3,400 (1,900)	69	99
C	Base	300 (224)	7	25 (19)	12	19,000 (10,500)	100	94
	Wall	100 (77)	4	15 (12)	7	6,800 (3,800)	68	98
	Column	50 (37)	4	15 (11)	3	3,400 (1,900)	78	98

Table 9.6: Performance estimates for Companies A and C

9.3.5 Comparison of simulation results

A comparison of the performance estimates (from *CBRSim*) for each company shall be made in this section. Initially this is done by examining the process performance measure estimates individually. The first measure that shall be examined is direct cost.

Direct Cost

Estimates of the direct costs (to the contractor) associated with the selection of each RMC supplier for the whole project are shown in table 9.7. Where project direct cost = (direct cost (base pour) x no. base pours) + (direct cost (wall pour) x no. wall pours) + (direct cost (column pour) x no. column pours). The values of direct cost (base, wall or column) are taken from table 9.5 for Company B and table 9.6 for Companies A and C.

Company B provides the smallest total cost estimate. The margin of difference between this company and the next cheapest (Company A) is of the order of $21,000 (£11,600). This represents a significant difference and unless Company B's estimate for duration is significantly larger than the other suppliers or the project is late (constraint 1 of section 9.2),

Company B would most likely be offered the contract on the basis of cost alone. The duration estimates associated with each supplier shall be examined next.

Company	Direct Cost (base pour) in $ (£) (1)	No. Base Pours (2)	Direct Cost (wall pour) in $ (£) (3)	No. Wall Pours (4)	Direct Cost (column pour) in $ (£) (5)	No. Column Pours (6)	Total Direct Cost in $ (£) (1 x 2) + (3 x 4) + (5 x 6)
A	18,200 (10,000)	34	6,700 (3,700)	40	3,400 (1,900)	36	1,007,000 (556,000)
B	18,300 (10,100)	34	6,200 (3,400)	40	3,200 (1,800)	36	985,000 (544,000)
C	19,000 (10,500)	34	6,800 (3,800)	40	3,400 (1,900)	36	1,040,000 (575,000)

Table 9.7: Total cost estimates by RMC supplier

Duration

Estimates of the project duration for each supplier are shown in table 9.8. With respect to duration, Company B outperforms the nearest competitor by a significant margin of 120 hours or the equivalent of 3 full weeks of construction work. Thus if either the project is late (constraint 1 of 9.2) or over budget (constraint 2 of 9.2) then Company B should be selected as supplier; they are cheaper and can deliver faster than the others.

Company	Duration (base pour) (hours) (1)	No. Base Pours (2)	Duration (wall pour) (hours) (3)	No. Wall Pours (4)	Duration (column pour) (hours) (5)	No. Column Pours (6)	Total Duration (hours) (1 x 2) + (3 x 4) + (5 x 6)
A	11	34	7	40	4	36	787
B	12	34	5	40	3	36	669
C	12	34	7	40	3	36	794

Table 9.8: Total duration estimates by RMC supplier

Use of space

A measure of how efficiently each supplier will use space in the project can be derived from the estimates of truck mixer utilisation; if truck mixer utilisation is high (approaching 100%) this signifies that there will be little queuing of truck mixers on the project site and hence space is being used efficiently. This factor would only affect the decision on which supplier

should be used in the case where work space is at a premium and the queuing of truck mixers is likely to impede the progress of other processes, otherwise known as project constraint 3 (of section 9.2).

Estimates of the truck mixer utilisation resultant from each supplier are shown in table 9.9. Company A would perform the RMC operations with the most efficient use of project site space, followed closely by Company C. Company B would perform operations which required more queuing of truck mixers than the others, but only by a small margin.

	Pour Type			
Company	Base	Wall	Column	Average
A	97	99	99	98
B	91	98	90	93
C	94	98	98	97

Table 9.9: Truck mixer utilisation estimates (%) by RMC supplier

9.3.6 RMC supplier selection

If this project were either late or over budget then it would be subject to constraints one or two (of section 9.2). In this instance Company B should be selected because this would result in a cost saving of $21,000 (£11,600) and a time saving of three-weeks from the services offered by the others. If, however, the site space was the main constraint of the project (constraint 3 of section 9.2) then Company A should be selected, as it would make the most efficient use of truck mixers as a higher utilisation rate translates into less queuing of the plant on the site.

9.4 Chapter conclusions

This chapter explored how CBRSim can be used in practice to aid a construction practitioner to make decisions regarding the RMC supply process. Two decisions were examined: how many truck mixers are required in an operation; and, who the supplier of RMC should be.

In both cases, it was shown that the performance of an operation could be planned to have a maximum impact on the main project constraint. For instance, in a project where cost was the main constraint a RMC operation could be planned so that it was performed at a minimum

cost. This planning would involve choosing the truck mixers that would result in a minimum cost (in the first decision above) or choosing the supplier that would afford the minimum cost (in the second).

In the decision regarding RMC supplier, one offered a flexible delivery rate of RMC and number of truck mixers to deliver the material. This flexibility created a competitive advantage over the other two suppliers; highlighting the impact that resource allocation can have on cost and duration of RMC operations.

CHAPTER 10

ESTIMATION OF OTHER CYCLICAL, STOCHASTIC PROCESSES

Work towards fulfilling the aims of this thesis, by means of considering the RMC process, is nearing its conclusion. However, it is important that the thesis aims of investigating *construction* processes, and not just RMC, are considered and this chapter discusses how the model might be extended to other areas of construction.

10.1 Introduction

It is hoped that the ability of *CBRSim* to provide accurate estimates of the performance of the cyclical, stochastic process of RMC supply can be replicated when other cyclical, stochastic processes are being modelled. If this replication is possible then *CBRSim* could potentially have a much larger impact on the performance of not only construction projects, but on the entire industry.

To investigate the generic modelling ability of *CBRSim* a model of the earthmoving process, which is both cyclical and stochastic, has been developed and validated. The validation of this model demonstrated that *CBRSim* has an ability to accurately estimate the earthmoving process, providing an indication that *CBRSim* can recreate other cyclical, stochastic processes.

Chapter 10 is structured as follows:

- **Section Two.** The earthmoving process is defined. The sources of data are discussed and the variables to be used in modelling are identified. Subsequently a model of the earthmoving process is developed and validated.
- **Section Three.** Two other cyclical, stochastic processes: pavement asphalting and structural steel erection were observed in this PhD project. These processes are defined and could potentially be recreated accurately using *CBRSim*.
- **Section Four.** Some conclusions are drawn.

10.2 Modelling the earthmoving process

The earthmoving process was observed in early 2004 at the Glasgow south orbital road/M77 extension project in Scotland. This project was not observed over a long enough period of time to collect a data set large enough for use in this study; rather the data used in this project was collected by Smith over the years 1992 to 1994. The details of the available data set are given in 10.2.2.

This project was commissioned under the UK Government's PPP scheme and was undertaken by the Connect Consortium (consisting of Balfour Beatty Capital Projects Ltd. and Atkins Investments). It is now complete and was worth 130 million pounds, involving the upgrade of 15km of existing roads to motorway status and the development of 9km of new motorway grade road.

10.2.1 Defining the earthmoving process

The earthmoving process is defined in figure 10.1. The principle resources involved are: a loader(s), which removes the earth from the ground; usually several haulers, which transport the material to a specified dumping area; and the employees to operate the plant involved.

The earthmoving process consists of the following activities:

- To begin, a hauler is positioned adjacent to the loader.
- Next, the loader fills up the hauler.
- The hauler then transports this load to the area designated to hold the material.
- The hauler dumps the material provided there are no other haulers queuing and there is some available space in which to dump the material.
- Subsequently the hauler returns to the loading area.
- As there are many haulers working within the system, the likelihood is that a queue of haulers may form at the loader – as only one hauler is usually loaded at a given time. This queuing is undoubtedly a source of frustration for a contractor, but is seen as a consequence of the process. Only when an especially large amount of time with haulers remaining static is experienced an attempt may be made to reduce the queuing of trucks.
- Finally, to complete the process a hauler is again positioned at the loader.

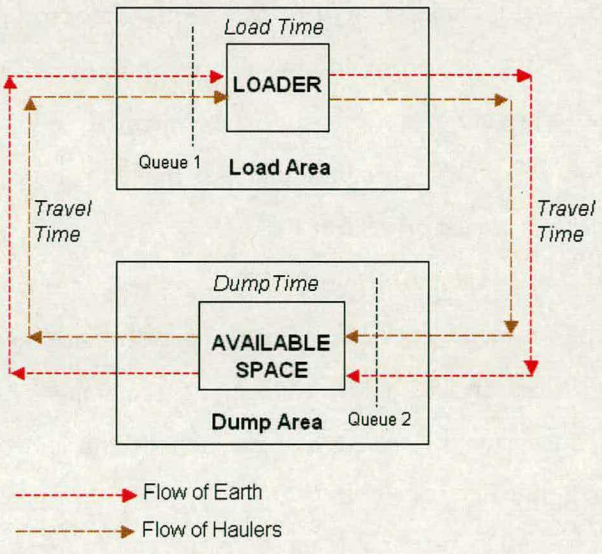


Figure 10.1: The earthmoving process

Now that the earthmoving process has been defined, the data available for modelling can be discussed.

10.2.2 Data available for modelling

The data available for this study was collected by Smith (1992 – 1994) during a time study of four road-building projects, undertaken by the same contractor in the UK. These projects varied widely in terms of: the quantity of soil that was required to be moved, from the relatively small A52 Ashbourne relief road (120, 000 m^3) to the M1/A1 link road (3,000,000 m^3); the location of the projects (South, Midlands and North of England); the soil types encountered; and numerous other factors which are difficult to determine. In total, 141 earthmoving operations were recorded, 90 of which were deemed suitable for use in this study; the others were discounted due to missing data. Two-thirds of these 90 operations were selected for use in model development with the remaining third being retained for use in model validation.

10.2.3 Variable selection

The model of the earthmoving process developed in this project was done so within the modelling framework devised in chapter 5 (figure 5.9). This section covers steps 1, 2 and 3 of the modelling framework which involve:

- Listing all available variables (from the collected data) and classifying if they would most likely be known by a construction practitioner at the time of ordering plant (haulers and loaders) in this case. This time represents the lead-time required for a supplier to provide the plant required and is roughly 4-5 weeks before the earthmoving operation is due to commence.
- From this list, all of the variables that are most likely known at the time of ordering plant are placed on a short-list.
- The third step involves performing a linear regression analysis on the short-list variables, with respect to each of the process performance measures (productivity, cost, hauler utilisation and loader utilisation). From these four linear regression analyses a consensus is formed as to which of the short-listed variables are statistically significant.

The steps detailed above were followed and the number of earthmoving variables was reduced from 19 to 3 (see figure 10.2). These 3 variables shall now be used to develop a *CBRSim* model of the earthmoving process.

10.2.4 CBRSim model development

To recap, *CBRSim* consists of two modules working together to form a hybrid model. A CBR module provides the probability distributions (and their parameters) that represent the times of interest of the earthmoving process, given a specific set of input variable values. Figure 10.1 shows that there are four 'times of interest' contained within the earthmoving process. They are, the load time (time taken to fill a hauler); travel to dump time; dump time (time taken to empty a hauler); and the travel to loader time. An example of the case base, which contains records of past earthmoving operations, is shown in figure 10.3.

The same indexing and retrieval mechanisms employed in the *CBRSim* model of chapter 8 have been employed in this model. It is these mechanisms which govern the solution provided by the CBR module.

When a new (target) case is presented to *CBRSim* by a construction practitioner, the (four) probability distributions (and their parameters) that represent the times of interest of the earthmoving process are provided by the CBR module to the DES module. The DES module works in the manner outlined in chapters 6 and 8 to recreate the earthmoving process, until some stopping condition is met. The stopping condition is a user-defined quantity of earth required to be moved. The *CBRSim* model then provides a summary of the process performance measures (e.g. productivity), as calculated from the model's statistical counters. It is these process performance measures that may prove useful to a construction practitioner, but only if the model is valid.

Variables	Determinability	Remaining Variables (short list)	Statistically Significant Variables (for use in modelling)
Haul road soil parameters	Medium	Haul road gradient	Hauler Manufacturer and Model
Haul road gradient	High	Hauler Manufacturer and Model	Number of haulers used
Haul road rolling resistances	Low	Loader bucket volume	Total volume to be moved
Haul road length	Medium	Bucket passes per load	
Hauler Manufacturer and Model	High	Number of haulers used	
Plant operator ability	Low	Total volume to be moved	
Load area characteristics	Medium	Month of excavation	
Type and load of obstructions	Low		
Weather	Low		
Material susceptability to weather	Low		
Load cycle times	Medium		
Hauler manoeuvre times	Medium		
Hauler travel times	Medium		
Load dump times	Medium		
Loader bucket volume	High		
Bucket passes per load	High		
Number of haulers used	High		
Total volume to be moved	High		
Month of excavation	High		

Figure 10.2: Variable selection made using the modelling framework (of chapter 5)

Case Number	Hauler Manufacturer and Model	Number of Haulers Used	Total Volume to be Moved (cum)	Load Time (minutes)			Travel to Dump Time (minutes)			Dump Time (minutes)			Travel to Loader Time (minutes)		
1	Caterpillar D400D	6	400	normal	15	5	lognormal	5	1	normal	1	1	lognormal	8	1
2	Volvo A25	12	12.6	lognormal	12	2	lognormal	8	2	lognormal	2	2	lognormal	8	1
3	Volvo A35	8	12.6	gamma	9	2	gamma	7	6	gamma	3	2	gamma	7	6
4	Caterpillar D400D	7	12.6	weibull	15	2	weibull	9	5	weibull	4	2	weibull	9	5
5	Volvo A25	10	12.6	uniform	8	10	uniform	12	3	uniform	2	1	uniform	12	5
6	Volvo A25	6	5.5	beta	12	1	beta	5	4	beta	3	1	beta	8	2
7	Caterpillar D400D	8	5.5	exponential	9	6	exponential	5	1	exponential	4	2	exponential	5	1
8	Volvo A35	14	5.5	normal	10	4	normal	6	4	normal	3	2	normal	7	3
9	Caterpillar D400D	3	10.9	beta	7	2	beta	7	2	beta	5	2	beta	8	1
10	Volvo A35	8	10.9	lognormal	8	5	normal	6	2	lognormal	3	2	normal	7	2

Figure 10.3: A sample from the case base of *CBRSim*

10.2.5 CBRSim model validation

As stated above 30 earthmoving operations were retained for use in the validation process. Each of these observed operations was recreated by the *CBRSim* model, resulting in estimates of the process performance measures, given the observed operating conditions (e.g. total volume to be moved, etc.). These estimates were compared with the observed process performance measures, over the entire validation data set. A graphical display of this

comparison is shown for productivity in figure 10.4. The 1:1 line indicatives a model that has provided accurate estimates of reality (observed values).

Figure 10.4 shows that the estimates produced by *CBRSim* of the earthmoving process productivity are accurate, as the results are all close to the 1:1 line. Over the whole validation set *CBRSim* was able to estimate the process productivity to within +/- 2% of the observed. The process performance measure that *CBRSim* had most difficulty in estimating was hauler utilisation that was within a range of +/- 9% over the validation set.

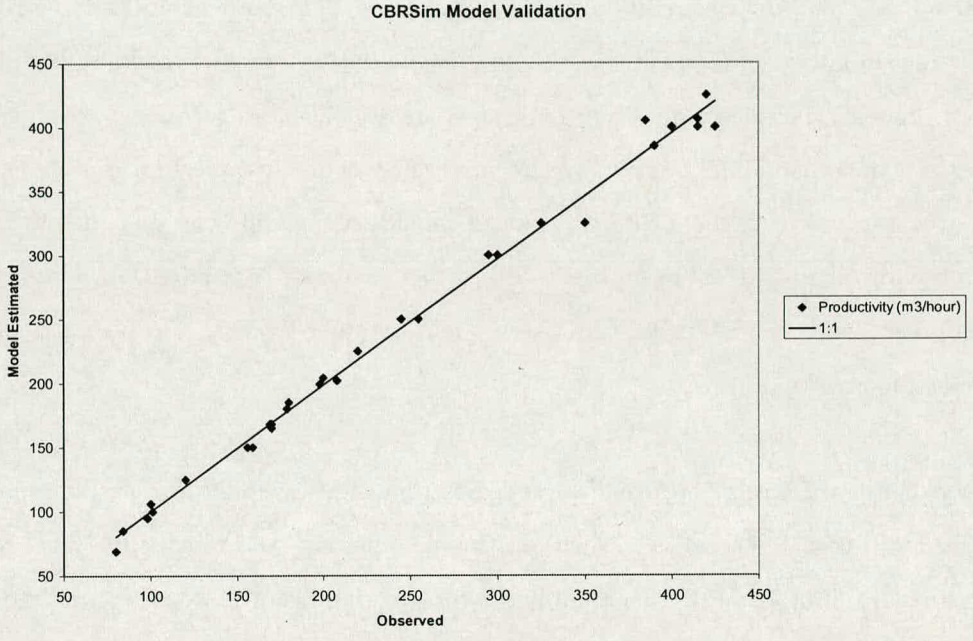


Figure 10.4: Correlated inspection results of *CBRSim* productivity estimates

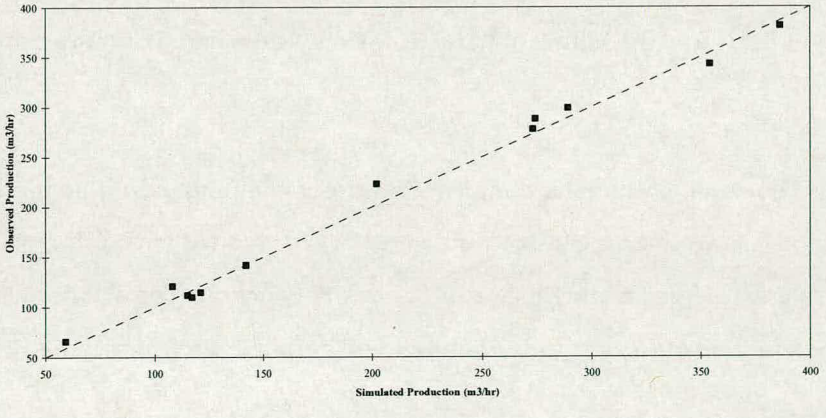


Figure 10.5: Correlated inspection results of simulation productivity estimates (Smith, 1995)

A comparison can be made between the results obtained from CBRSim and a simulation model, Trucksim, developed by Smith in 1995 (shown in figure 10.5). Although a different amount of data was used in the validation of the two models it is still possible to compare the fit of a 1:1 line to the correlated inspection validation. There is a small, but noticeable, improvement in the accuracy of the validation results from the CBRSim model.

A further comparison between the *CBRSim* models of the RMC supply process and the earthmoving process is possible. The CBRSim model of the RMC supply process had a best estimate of +/- 3% (for productivity) and +/-20% (for on-site resources utilisation). The above results revealed that *CBRSim* is more able to recreate the earthmoving process than the RMC supply process. The question within this chapter is 'can CBRSim recreate cyclical, stochastic processes other than RMC supply?' A tentative answer to this question is yes. The above comparison is indicative that *CBRSim* is indeed capable of modelling at least one other process (earthmoving) than RMC supply. However, further evidence is required, in the form of the development and validation of *CBRSim* models of other cyclical, stochastic construction processes, to confirm this.

Other cyclical, stochastic construction processes (than RMC supply and earthmoving) were observed in this PhD project, such as asphalt paving and structural steel erection. Unfortunately, they have not been studied over the significant periods of time required to build up a sufficient data set upon which to develop a model. Rather, these processes were observed for a sufficient length of time to allow a definition of each process to be made, which may be useful in the future modelling of these processes. These definitions are provided in the next section.

10.3 Further cyclical, stochastic construction processes that could be modelled

Two further cyclical, stochastic construction processes, pavement asphalting and structural steel erection, were observed during the course of this PhD project. The details of the observed projects and a definition of each process are provided in the subsequent sections.

10.3.1 Asphalt paving process

The process of paving a road with asphalt was observed during the construction of the M1/A1 Sprucefield Interchange in 2003 in Northern Ireland. A map of how the process would be viewed in *CBRSim* is shown in figure 10.6.

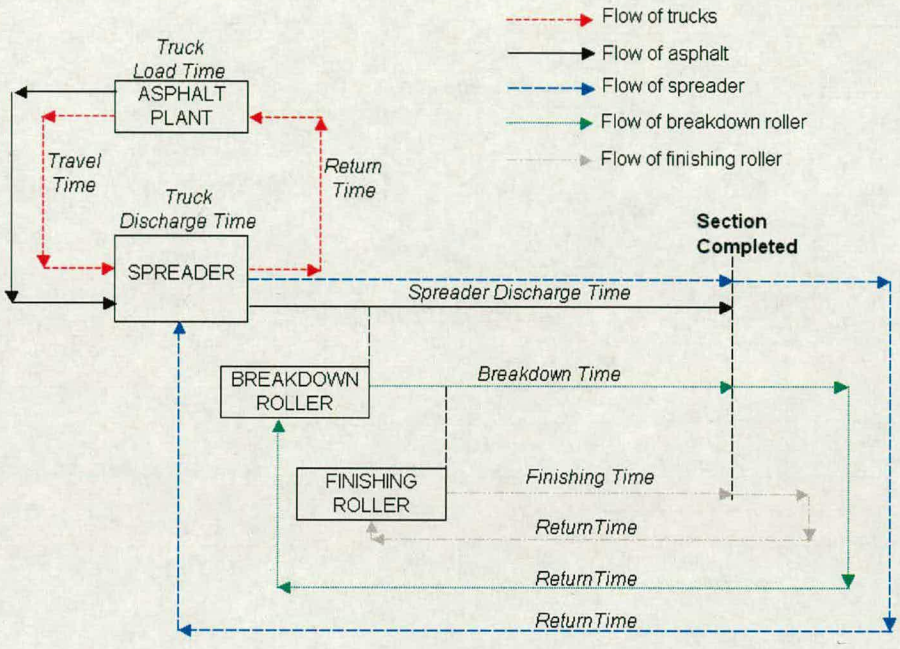


Figure 10.6: Asphalt paving process

10.3.2 Structural steel erection

The process of structural steel erection is another stochastic, cyclical process that could be estimated by *CBRSim*. This process was observed during the construction of Stanley Hall, on the campus of UC Berkeley, CA, U.S.A in Spring 2005. A map of how the process would be viewed in *CBRSim* is shown in figure 10.7.

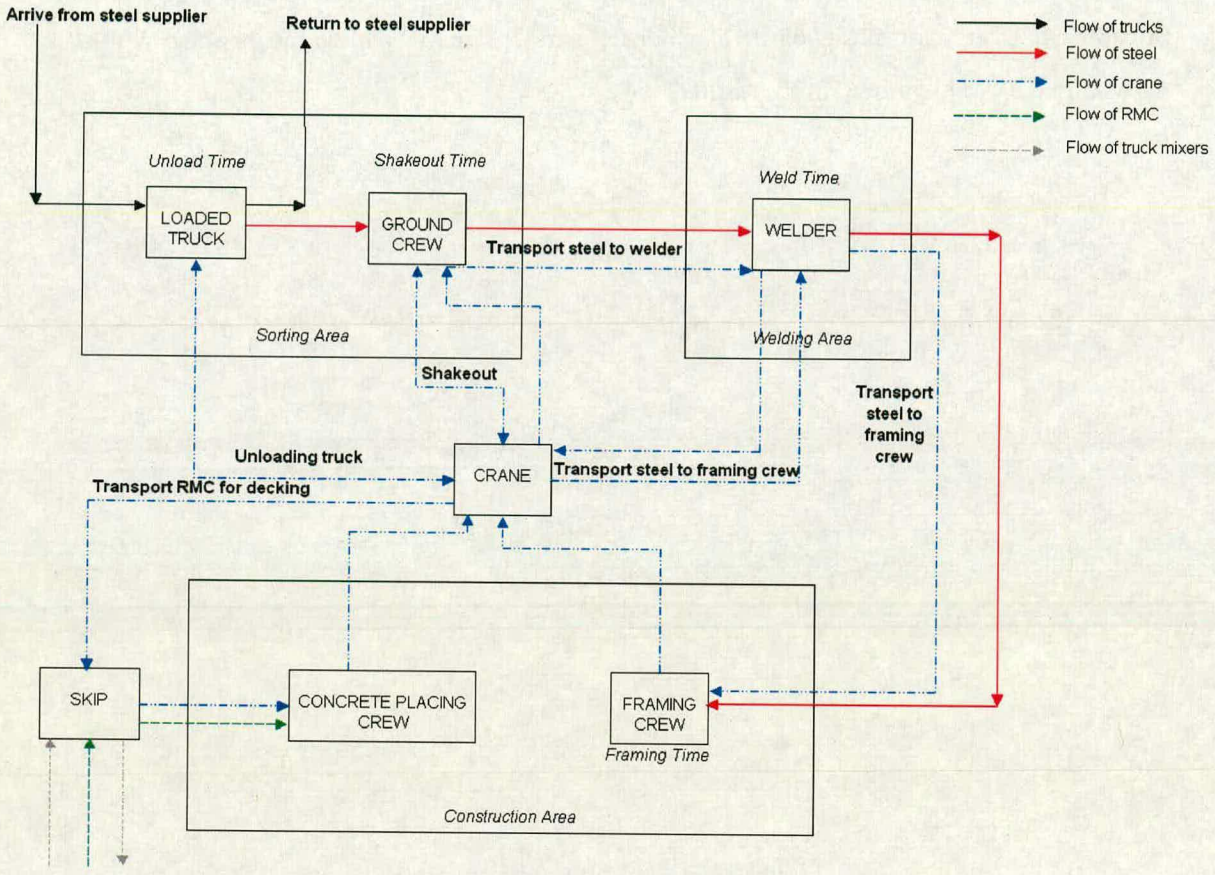


Figure 10.7: Structural steel erection process

10.4 Chapter conclusions

It was found that CBRSim was capable of estimating the earthmoving process more accurately and reliably than the RMC supply process. This could be due to the earthmoving process being more linear than RMC supply and one reason for this may be that earth haulers rarely leave the confines of a construction site, whereas RMC truck mixers often travel on public roads and encounter variable levels of traffic.

As CBRSim was found to be capable of estimating the earthmoving process accurately and reliably, this points to the ability of the model to recreate cyclical, stochastic processes in general.

CHAPTER 11

CONCLUSIONS AND FUTURE WORK

At the start of this thesis, in chapter 2, three main aims were set down. Now, at the end of this thesis, the ways in which these aims have been met is presented. In many cases, the research has gone beyond these aims and this chapter will point out where. Finally, it is important that suggested paths for futher work are indicated.

11.1 Introduction

This chapter contains the conclusions that may be drawn from the work presented in this PhD thesis. Where appropriate, a summary of the findings of this research that provide evidence to support a particular conclusion are included.

Chapter 11 is structured as follows:

- **Section Two.** A summary is presented of the problem being addressed by this research project.
- **Section Three.** Conclusions are drawn from the processes of data collection and pre-modelling data analysis that were undertaken in this project.
- **Section Four.** The processes of model development, evolution and validation are summarised and conclusions are drawn.
- **Section Five.** A summary of how the model may be used is provided and conclusions are drawn as to how this model may be implemented in practice.
- **Section Six.** The limitations of the work presented in this research are expressed and some suggestions for future work are presented.
- **Section Seven.** Personal learning outcomes from the process of undertaking this research project are presented.

11.2 Summary of the original problem and problem evolution

Construction projects frequently overrun and finish over budget. This is in part due to the lack of control that construction practitioners – planners and managers – have over the construction schedule at a process level. Contemporary planning tools, such as critical path method (CPM) are adequate for planning at a project level, but are not of sufficient detail to help a practitioner to allocate resources to a process to maximise the performance. Thus the aim of this project was to develop a practical computer-based model to enable practitioners to plan projects at a process level and to improve their projects. This research was focused upon a specific type of construction process – those that are stochastic and cyclical, because such processes are: difficult to predict and hence control; and they are widespread throughout construction projects. Examples of processes that are stochastic and cyclical are crane operations, formwork erection and scaffold erection. Initially, a focus was placed on the process of ready-mixed concrete (RMC) supply.

To re-iterate the three main aims of this thesis were:

- To investigate methods to improve the performance of construction projects and the UK construction industry.
- To develop a model capable of practical application.
- To learn more about stochastic and cyclical construction processes.

An initial period of data gathering and analysis confirmed that only a non-linear model would be sufficient to recreate these processes. Most of the previous models of cyclical, stochastic construction processes had been simulation models, and this methodology was used as a basis for this research project. It was soon discovered that simulation could be a useful tool in allowing experiments to be performed with different quantities of resources – helping to form a picture of the optimum resource allocation for a particular operation of a process. However, the idea of simulation modelling had not become widespread in industry despite these experimental capabilities. This is attributed to the fact that a simulation model cannot provide accurate estimates of process performance, unless the model is first provided with an accurate reflection of the statistical nature of the process – this became known as *the complexity problem* in this research project. This problem was encountered in this research project and led to an evolution in the project's aims. The new aim was to continue developing a model that accurately recreates cyclical, stochastic construction processes and solve *the complexity problem* at the same time. No longer was just an accurate model of construction processes being sought; rather a solution to one of the fundamental drawbacks of simulation modelling was also.

11.3 Data availability and pre-modelling analysis

The premise of this research project was to develop models that are both pertinent to, and practical in, an industrial setting. This premise sits well with the common philosophy of modelling that putting "rubbish" into a model results in receiving "rubbish" out of a model, because a model that does not produce an acceptable representation of the problem at hand will be neither pertinent to or practical for industry. To this end, it was determined that any models produced in this research were to be based upon data collected from real construction projects.

- The details of the data collection in this research project may be summarised as follows:
- A substantial data set was initially available for use in this project, having been collected through observation of the RMC supply and use processes by another PhD student, Dunlop (2000). This data was collected from the construction of four wastewater treatment plants in various locations around Scotland, and consisted of 232 observations and 28 variables. While this data set was and still is acknowledged to be extremely comprehensive, particularly in the construction field, it was felt that there were some deficiencies that needed to be improved upon. Specifically, concerns were raised regarding: the impact that the project location may have on the data set; the effect of the type of project on the data set; and the impact of the weather on the data set.
- Project location was a matter of concern, because the observed projects took place in close proximity to one another, and there may have been a possibility that the models developed from this data would have been only applicable for projects in similar locations to those of the original data. In the same vein, the observed projects all concerned the construction of wastewater treatment plants, raising the possibility that the models produced in this research would only be applicable to this type of structure.
- To counteract the above concerns, an additional set of data was acquired. This data set of 50 pours was collected by a construction contractor during observations of the construction of a motorway viaduct in the North of England; thus, providing data that is from a different location and type of project.
- The literature survey performed in this research project brought to the fore the need to include the influence of weather on the observed pours. This need was fulfilled through the acquisition of historical data from the UK Met Office. The historical data resulted in the number of variables contained in the data set increasing from 28 to 34 (of which 26 are independent).

The above data set provided an opportunity to develop accurate models of the RMC supply process. This opportunity was seized by analysing the data with a view to reducing the number of modelling options considered to be suitable in this research, thereby providing direction whilst improving the efficiency of the research itself. This reduction was performed

in two ways: to determine if standard statistical analysis techniques (e.g. linear LS regression) were appropriate for use with the data; and, to determine if a non-linear model of the process was strictly necessary. Knowledge of the former is important because a number of modelling methods are based upon statistical methods or at least statistical information, and thus an inability to use standard statistical methods would have a significant impact on the type of modelling method that could have been used in this research. The latter is pertinent because this project had at its foundation the principle of parsimony, and thus if a simple, linear-type model would provide a sufficiently accurate model of the process then there would be no requirement to complicate matters by developing more complex non-linear models of said process.

The details of the above investigations may be summarised as follows:

- Two methods were used to determine whether standard statistical methods fitted with the project data. One was to establish the distribution (e.g. normal) of the data. This is because standard statistical measurements (e.g. correlation analysis or linear LS regression) cannot be utilised with data which is non-normally distributed. The project data was found to be normally distributed. The other method was to investigate if the data set contained outliers.
- Outliers are observations (or subset of observations) that appear to be inconsistent with the remainder of that set of data. The presence of outliers can adversely affect the ability of a model to recreate a process and were thus of interest in this research. If a small number of outliers were present in a data set, then it may be possible to remove these outliers. However, a data set rife with outliers would be unable to utilise standard statistical techniques, and a set of robust statistical practices would be required instead.
- A new method of identifying outliers in data was proposed, developed and validated in this thesis. This method was thought to be required because the current common method utilises the Mahalanobis distance measure as a basis for defining outliers, and this measure requires the calculation of the correlation variance (covariance) matrix. This particular matrix is thought as unreliable when dealing with small data sets (less than 100 observations), and as construction data sets are often small an alternative method of identifying outliers, not based upon

the calculation of the covariance matrix was deemed desirable. Thus, a measure based upon the more simple Penrose distance measurement was presented in this thesis. The method was validated on an independent set of data and was found to out-perform the established method. Thus, it was used on the project data set. However, no outliers were found.

- Taking into account that the data was normally distributed and contained no outliers, it was declared that standard statistical methods were applicable in this work. An example of one such method used in this project is an examination of the correlation between the explanatory variables of the project data, with a view that if there are two highly correlated variables one may be removed without any significant loss of information. This analysis resulted in two sets of two variables being found to be highly correlated, and afforded to removal of two variables, thus reducing the data set from 26 explanatory variables to 24.
- The investigation of whether the model needed to be linear or non-linear was initially conducted by fitting a linear LS regression equation to the project data. The findings of this study were that there were possibly some linear trends in the data, but there was no conclusive evidence. Further investigations took place, involving plots of the models errors (residuals) against the model's estimates of the dependent variables. If a model were to be (strictly) linear these plots would have shown no error for each estimate – this was not the case, and thus it was determined that a non-linear modelling method would be required in this research.

The data set available for use in this project proved indeed to be substantial, particularly in the context of construction research. While a data set of this magnitude provided a sufficient number of examples of concreting pours on which a model could be based, and this was of course beneficial to the project, there were some drawbacks caused by this quantity of data. For instance, it was determined by a combination of discussion with construction practitioners and intuition that a model based upon 24 explanatory variables would not be practical in an industry where time is as valuable as money. In essence, it was concluded that a model does not become practical and pertinent to industry solely because it is based upon real data; rather the practical, pertinent model contains a refined set of variables – based upon what a

practitioner can reasonably know at specific stages of a project and what is statistically and intuitively important.

A five-stage modelling framework was devised in conjunction with industry practitioners to help achieve a refined set of variables as discussed above. This framework involves deciding on what variables are likely to be known at the time of ordering plant or materials and then using linear regression to reduce this set of variables to a small, statistically significant group on which a model can be based to allow a practitioner to recreate the process, given a particular set of conditions, with ease. In this research the framework was used to reduce the number of explanatory variables from 24 to 6. Additionally, the modelling framework could be used in other projects as a basis of ensuring only the most relevant variables are considered for use in modelling.

At the end of this data collection and pre-modelling analysis process it was felt that a sufficient set of data had been gathered and a group of variables had been defined that would permit models to be developed that would prove both practical and pertinent in an industrial setting. This is because: the variables have been reduced to a small number that should be known by practitioners at the time of ordering plant or materials and are statistically significant; and the data on which the models will be based is taken from real projects, helping to substantiate the findings of this research in the eyes of construction practitioners and researchers.

11.4 Model development, evolution and validation

The three project aims stated in chapter 2 all had one central theme: a valid model of cyclical, stochastic construction processes would be developed. To reiterate the findings of the pre-modelling analysis process, this model must be non-linear. Through the consideration of previous research in this field, it was determined that a common method of modelling cyclical, stochastic processes should be initially considered. This method is discrete-event simulation (DES) and a DES model was to act as a basis for comparison for all subsequent models.

DES is well known for providing a user with an ability to experiment with a process, to manipulate the variables of that process in the search of an optimum combination of these variables. This experimental capability was thought to be applicable in this research work, as a planning engineer has a number of resources available (or can order a number of resources)

and these must be allocated to the performance of each process in a manner that is best for that particular project. Thus a practitioner can manipulate the possible combinations of resources in a DES model and examine the consequences with little effort and wasted time.

The details of the development of a DES model of the RMC supply process can be summarised as follows:

- The RMC supply process has 6 *times of interest* (e.g. travel time to project site) that when considered cumulatively describe the activity of a concrete truck mixer as it completes a whole cycle (e.g. batch plant to project site to batch plant) in the process.
- A probability distribution was used to represent each of these *times of interest*, and the DES model takes the average of 1000 estimates from each distribution to estimate the cycle time of one truck mixer. Thus, the probability distributions are most important in a DES model; an inaccurate distribution will provide "rubbish" into the model, resulting in "rubbish" out.
- While DES is undoubtedly a useful tool, as it allows experiments to be performed with ease, the actual results from these experiments are of dubious integrity. This is because a change in one of the variables will result in a change (even a slight one) in the shape of the distributions that describe the *times of interest*. Thus to perform an experiment accurately not only should the value of a variable change, but the shape of distribution in the model should be adjusted accordingly. This is not to say that the experiments performed by DES models are not of use; they are – as the trends shown in these experiments are most likely to provide a true reflection of reality. However, a mere reflection of reality is not what was required in this work; there is a need for accurate estimates of the process productivity, given changes in the model variables, not just the trend that the productivity goes up if a particular variable is altered in a particular manner.
- The above point is illustrated well in the validation process of the original DES model (as discussed in chapter 6). It was found that this model provided promising results, but changing the distributions that represent the *times of interest* and drive the simulation model was difficult. The problem of how distributions change with changes in the model's variables became the central focus of the

thesis, as simulation was deemed useful, but only if the estimates of the process performance were accurate. This problem became known as *the complexity problem* and a solution was sought in this research project.

With the identification of *the complexity problem,* came the evolution of this research project's remit; no longer was the aim to just develop a model, rather it was to improve a deficiency which has blighted simulation modelling since its inception.

The solution came from the field of Artificial Intelligence (AI) in the form of case-based reasoning (CBR). CBR works by using solutions to similar, previous problems to solve a new issue, and was thought to be capable of suggesting the shape of the probability distributions that represent the *times of interest* of cyclical, stochastic construction processes and how the shapes of these distributions change under different conditions.

An investigation of how CBR worked with construction data was undertaken concurrently with the standard practice in CBR modelling of experimentally determining the optimum make-up of the model. The make-up of a model in this instance refers to the mechanisms: to store past examples (cases) and their solutions in a standardised manner in a database (case base), through a process known as case indexing; to retrieve the most appropriate solution from the available past examples (cases) to be utilised as a solution to the current problem, in a process known as case retrieval. To ensure the modelling method is as practical as possible, the speed of retrieving an acceptable solution is a crucial factor, and was considered in the search for an optimum model structure.

Indexing is a useful element of a CBR model, because it ensures: that the information is stored in a standard form – reducing the potential for the information to be of a poor quality; and indexing reduces the amount of information that must be searched. For example, the un-indexed variable pour volume would result in a search of any number to one decimal place (e.g. 10.1 m^3) from between 8m^3 and 400 m^3. Whereas, breaking this variable into ten crisp indices would result in a search of integers between 1 and 10 – clearly a much more efficient method of retrieving information, which is an important factor in this project.

The following provides a summary of the experiments that were conducted to find the optimum combination of index and retrieval mechanisms for use with construction-type data:

- Indexing mechanisms. Three types of indexing mechanisms were considered in this research project: one which categorised the project data into crisp indices, where each index was definitely different from another; one which categorised data into fuzzy indices; and one which did nothing to the data – the data was stored in the case base in a raw form.
- Three methods of defining the size of crisp indices were considered because there is no generally accepted method of performing this task. These three methods were taken from the categorisation methods used in the formation of histograms, thus providing an element of mathematical rigour to the task.
- Fuzzy indexing of the project data was considered a necessity because it was difficult to accept that the data differed substantially between the distinct indices – as is assumed in crisp indexing. The problem is illustrated well by the categorisation (indexing) of the data for the variable pour volume. Using a crisp indexing method resulted in a pour volume of 79.3 m^3 being placed in a separate category to a pour of 79.4 m^3; this was difficult to swallow. The fuzzy indexing of this information increases the reality of the CBR model, allowing for 'grey' areas in the indices, rather than the 'black and white' view of crisp indexing.
- There is no agreed method for how to successfully define fuzzy indices, but what is agreed upon in the relevant literature is that the sizes of fuzzy indices are crucial to making the method work. Thus, a conservative method of defining fuzzy indices was considered in this project. This method involved choosing the original width of the fuzzy indices and then iteratively adjusting these widths by +/- 10% up to a change of +/- 40% from the original. This resulted in an enormous number of models to compare, hence the conservativeness of the method.
- Retrieval mechanisms. Case retrieval usually involves making a measurement of the mathematical distance between the case to be solved (target case) and each case in the case base; a measurement known as similarity. There are numerous ways of performing such a calculation, thus a similar approach to indexing was adopted in that a number of different methods of similarity calculation were considered. There was however a difference in the methods that could be used for

the models that had crisp and fuzzy indexed cases: crisp models used a simple mathematical formula to calculate similarity; whereas fuzzy models had a more complex calculation of similarity to perform.

- Regardless of which retrieval mechanism was being considered, there were two versions: one which added weightings to the variables, to add a sense of reality to the models; and, one which left the variables un-weighted.
- The fuzzy models were shown to be more accurate than some of the crisp models, and could prove to be a useful technique of increasing the reality of a model. However, the choice of fuzzy index size is the crux of the matter and a more rigorous method of defining this should be investigated.
- It was found that on average the crisp and fuzzy models takes 600 seconds to produce an answer. This time of 10 minutes was thought to be prohibitive to the acceptance of the method by industry; a thought later confirmed during informal discussions on the matter with construction managers at MJ Gleeson in Scotland and Gayle Manufacturing in California. Thus a more efficient modelling method was sought in this project.
- The solution to the efficiency problem was adapted from the method of identifying outliers devised in this project (and detailed in chapter 5). To reiterate this method, the data is combined into clusters using a standard mathematical clustering technique and then the distance between each data point and the centre of each cluster is measured using the Penrose distance measurement. Any data points that have a large Penrose distance (relative to the others) from each of the clusters would be declared an outlier.
- The cluster-based retrieval method works by clustering the cases into a specified number of clusters (usually 3 or 4). The Penrose distance measurement is then used to calculate the distance between the target case and each cluster. The nearest cluster to the target case, defined by having the minimum Penrose distance, is then focussed upon. This cluster is likely to contain a much smaller number of cases than the original case-base, resulting in a greatly reduced search of the cases. The simple mathematical similarity measure utilised in crisp models is then used to measure the similarity between the target case and each case in the particular cluster.

- It was found that this method of case retrieval takes on average 120 seconds to provide a solution. Thus, a saving of 8 minutes from both the crisp and fuzzy models is plausible using the cluster-based method. This substantial increase in efficiency could result in a model that is more acceptable to construction practitioners due to the time-savings made over the alternative considered methods.
- All of the CBR models were validated against the motorway viaduct data and it was found that the weighted-versions of the models were more accurate than non-weighted. The optimum model was found to be that which utilised the cluster-based retrieval mechanism. The increase in accuracy is likely to be due to the fact that the most dissimilar cases held in the case base are discounted at an early stage in the retrieval process. This validation result was extremely satisfying, as not only is the cluster-based model the most accurate, but is additionally the most efficient.

The above study proved that CBR was capable of dealing with construction data and hence could prove useful in solving *the complexity problem* of simulation modelling. In the investigation of whether CBR would indeed provide a solution to this problem a hybrid model was developed.

The above hybrid model consisted of: a CBR model, based upon the optimum make-up discussed above, that was used to suggest the probability distributions that represent the *times of interest* of the RMC supply process; and a DES model that utilises these distributions to simulate cycles of the concrete truck mixers in the process, to provide estimates of how the resources in the process will perform.

The hybrid model was named *CBRSim* and in a validation study, was proved capable of estimating the productivity of the RMC supply process to within +/- 3% accuracy from the validation data, which is an acceptable level of exactness. From this result, it is possible to declare that some of the aims of this research project had been fulfilled: a solution to the complexity problem had been found; and a practical model has been developed. This can be declared because the model is founded upon real data, involves the user inputting information for the variables that are likely to be known at the planning stage of a project, and a complete

set of experiments may be performed to a high degree of accuracy within four minutes, providing a number of options for a construction practitioner to consider.

CBRSim was shown to be capable of modelling the RMC supply process with a good degree of accuracy. The ability to recreate one cyclical, stochastic process was not the complete objective of the modelling methodology; rather it was to be generic and thus possess a capability for recreating any cyclical, stochastic construction process – widening both its applicability and appeal to practitioners.

The generic capability of the model was investigated by developing and validating a model of another cyclical, stochastic construction process – that of earthmoving. *CBRSim* was found to produce more accurate estimates of the process performance measures such as productivity of the earthmoving process, than it did for the RMC supply process. Thus, *CBRSim* was shown to possess some ability to model other construction processes.

11.5 Use of the model

Two methods of using the *CBRSim* model were explored in chapter 9. This demonstrated how the model may be used to help a construction practitioner reach decisions on fundamental areas of construction projects, such as how many resources should be allocated to a particular operation of a process and who should be appointed the supplier of materials to a project.

Current critical path method-style project planning is not sufficiently detailed to help a construction practitioner to plan and control construction at a process level; it is particularly unhelpful in the resource allocation problem as the resources available and the impact of those resources are not considered; similarly, the available suppliers of materials are not considered in critical path-style planning, thereby not providing any support to reach this decision. Resources are currently being allocated based upon the intuition of a practitioner. While the practitioner often has vast experience it is difficult to account for the stochastic behaviour of the processes for which decisions are being made, resulting in an allocation of too little or too many resources and a poor process performance. This means that key activities (work-tasks) are not being performed with a maximum efficiency, at a significant cost, in terms of time and money, to resource suppliers, sub-contractors, general contractors and ultimately clients.

CBRSim was shown to be useful in helping a construction practitioner reach decisions as to how many resources should be allocated. *CBRSim* is not intended to replace the human element of the construction industry, but merely provide support to humans by factoring into the analysis the stochastic behaviour of construction processes. The inclusion of this complex behaviour should result in a much more accurate picture being formed in the mind of construction practitioners as to how the process will react to adjustments in plant, materials and personnel.

Experiments can be performed quite readily using a simulation model. *CBRSim* is a simulation that can perform experiments that not only reflect accurately the trends of impacts that change has on process performance, as a normal simulation model, but the actual impacts of change on process performance itself. Thus, *CBRSim* can be used to perform experiments that permit a comparative analysis, of say 3 competing RMC suppliers, to be undertaken, because the actual impact that each of these suppliers would have on the process performance can be estimated with a fair degree of accuracy. To summarise the findings of chapter 9, a distinct advantage can be gained in terms of money and time through a simple analysis of the terms and conditions offered by competing materials suppliers.

It is envisaged that this type of analysis is readily transferable to other areas of construction project planning, such as the issue of selecting equipment suppliers.

11.6 Limitations of this research work and recommendations for future work

The research undertaken in this project is not all-encompassing and has limitations. The following provides a summary of these limitations:

- The processes modelled in this research project are only of a stochastic, cyclical nature. While these processes can constitute a significant proportion of a construction project and are thus important, other types of process that are performed in construction projects are not considered.
- The models developed in this project are based upon a limited collection of data. While significant effort has been made to ensure the data is unbiased, it is difficult to say with absolute certainty that this is the case. The risk that the model was based on a biased data set, which would result in a model only fit for use with

that particular data set, was minimised by the performance of a robust statistical data analysis.

- The models developed in this research project do not take into account issues which are more traditionally considered in research which has adopted a 'softer' approach. These are issues such as what drives people and what motivates them to improve their productivity.
- Case-based reasoning, the modelling methodology considered in this research project, is based upon the heuristic determination of the model parts (e.g. indexing). Although a significant effort was expended in the model development process, it was impossible to be exhaustive. Thus, there is a limit to the amount of types and combinations of types of the parts of a CBR model.
- This project did not take the validation process to the point where the *CBRSim* modelling methodology was implemented in a practical situation.

Some subjects of future work in this research area are:

- The size of the case bases and validation data sets should be increased in the existing *CBRSim* models of the RMC supply and earthmoving processes. This could be achieved through further observation of construction processes or through the collection of data from a construction contractor's records.
- Further observations and data collection should be made from other cyclical, stochastic process such as the asphalt paving process. Upon the collection of this data, further valid CBRSim models – based on this data – should be developed. This will allow a more rigorous examination of the CBRSim modelling methodology's ability to recreate cyclical, stochastic construction processes in general.
- Examine further the methods of indexing and retrieving cases from the *CBRSim* case base. Specifically efforts should be made to:
  - Examine methods of choosing the optimum crisp index sizes.
  - Optimise the iterative method of calculating the fuzzy index sizes, in a manner that reduces the effort required to perform the task; thereby making the method more appropriate for practitioners to use.

- Examine techniques of automating or at least reducing the effort to optimise the above methods of defining index sizes. For instance some of the available techniques that could be utilised are genetic algorithms and linear programming.

- The existing *CBRSim* models should be validated further through a variety of methods. For example, a group of construction practitioners could perform a formal assessment of the modelling methodology.
- An important piece of work that could be performed in the future is to further examine and define the pitfalls for modelling using *CBRSim*. This work should help to ensure that the construction practitioners using the model are doing so with an explicit awareness of the limitations of the modelling methodology, hence improving its usefulness in practice.
- Some work should be performed on proving the CBRSim modelling methodology to industry through demonstrations, conferences and trade-fairs.
- The *CBRSim* modelling methodology should be implemented in a practical, live situation. This would be the ultimate test, as to its applicability in industry.

11.7 Learning outcomes from this project

There have been many lessons learned during the course of this project; they range from the general, where a transferable skill or knowledge has been learned, to the specific, where knowledge or skills that are uniquely applicable to this research area have been learned. The following is a description of these skills and knowledge, beginning with those of a general-type:

- Project definition. I learned to define a project's boundary through carrying out a rigorous literature survey and review, discussing issues with professional people, using existing information and expanding upon that information through the collection of data.
- Modelling and analysis. I learned to state a hypothesis and then produce models in a robust manner to prove/disprove the hypothesis. After this, I used these models to solve problems.
- Disseminating information. I learned, from an early stage, to disseminate the findings of my research work through a variety of methods. These methods included writing journal and conference papers; delivering presentations at

conferences; giving seminars; discussing my research with a group of non-engineering students at UK Grad school. Through presenting my research to a range of people I learned to think deeply about whom the recipient of the talk, presentation, seminar, or paper would be, and the method of presenting the information in a way that will be inspiring and understandable to the audience.

The following is a description of the specific-type learning outcomes:

- Lean construction was initially thought to be closely related to the work undertaken in this thesis. This thought was changed during my time as a visiting scholar to the production systems laboratory at UC Berkeley, where I participated in the lean research group meetings. It became apparent during these meetings and outside study of the relevant literature that current lean construction thinking was tending towards a less technologically-based solution – where the events of the previous week in a construction project provided a basis for estimating the productivity for the coming week in the project (under the Last Planner system). Thus current lean construction thinking is diverging from the methods proposed in this thesis and lean did not become integral to this project, as was first thought.
- At the beginning of the project an examination of the supply chain for the materials, plant, etc., that were used in the construction processes under study in this project, was not considered as necessary. Again through my time at UC Berkeley the importance of supply chain management became apparent and an investigation into reducing the costs of concrete for contractors through vertically integrating the resources supply chains was made, and found to be capable of producing significant cost savings.
- Significantly, it was learned at an early stage that simulation modelling has a deficiency which prevents its widespread adoption in industry. This deficiency is the requirement for a practitioner to know relatively well the statistics which underlie a process and use that knowledge in simulation model development. The solution to this problem became a major part of this research project, and it was further learned that a suitable solution to this problem did exist.

REFERENCES

1. Aamodt, A., and Plaza, E. (1994). "Case-based reasoning: Foundational issues, methodological variations, and system approaches." *AI Communications*, 7(i), 39-59.
2. Abourizk, S. M., Halpin, D. W., and Wilson, J. R. (1991). "Visual interactive fitting of beta distributions." *J. Constr. Engrg. and Mgmt., ASCE*, 117 (4), 589-605.
3. Abourizk S. M., Halpin D. W., and Wilson J. R. (1994). "Fitting beta distributions based on sample data." *J. Constr. Engrg. and Mgmt., ASCE*, 120 (2), 288-305.
4. Abourizk, S. M., Knowles, P., and Hermann, U. R. (2001). "Estimating labor production rates for industrial construction activities." *J. Constr. Eng. and Mgmt., ASCE*, 127(6), 501-511.
5. Alarcon, L. F. (1993). "Modeling waste and performance in construction." *Proc., 1st Annual Conf. of the Int. Group for Lean Construction*, Espoo, Finland.
6. Alarcon, L. F. (1994). "Tools for the identification and reduction of waste in construction projects." *Proc., 2nd Annual Conf. of the Int. Group for Lean Construction*, Santiago, Chile.
7. Ammar, M. A., and Mohieldin, Y. A. (2002). "Resource constrained project scheduling using simulation." *Constr. Mgmt. and Econ.*, 20 (4), 323-330.
8. Anscombe, F. J. (1960). "Rejection of Outliers". *Technometrics*, 2, pp. 123-147.
9. Anscombe, F. J., and Tukey, J. W. (1963). "The examination and analysis of residuals". *Technometrics*, 5, pp. 141-160.
10. Anson, M., Tang, S. L., and Ying, K. C. (2002). "Measurement of the performance of ready mixed concreting resources as data for system simulation." *Constr. Mgmt. and Econ.*, 20 (3), 237-250.
11. Arditi, D., and Tokdemir, O. B. (1999). "Comparison of case-based reasoning and artificial neural networks." *J. Comp. in Civil. Engrg., ASCE*, 13 (3), 162-169.
12. B.A.A. (2005). "BAA-Heathrow: About Heathrow: Terminal 5." http://www.baa.co.uk/main/airports/heathrow/about_heathrow/terminal_5_frame.html (Accessed: 02/06/2005).
13. Ballard, G. (1994). "Lean construction and EPC performance improvement." *Proc., 2nd Annual Conf. of the Int. Group for Lean Construction*, Santiago, Chile.
14. Ballard, G. (2000). "Lean project delivery system." *White Paper #8, Lean Construction Institute*, California.
15. Ballard, G., and Howell, G. (1994). "Implementing lean construction: Reducing inflow variation." *Proc., 2nd Annual Conf. of the Int. Group for Lean Construction*, Santiago, Chile.
16. Ballard, G., and Howell, G. (1997). "*Shielding production from uncertainty: First step in an improvement strategy*." Tech. Rep. 97-1, Construction Engrg. And Mgmt. Program, Civil and Environmental Engrg. Dept., University of California, Berkeley, California.
17. Ballard, G., and Howell, G. (1998). "Shielding production: Essential step in production control." *J. Constr. Engrg. and Mgmt., ASCE*, 124 (1), 11-17.
18. Ballard, G., and Howell, G. (2003). "Competing construction management paradigms." *Proc. ASCE Construction Congress*, Honolulu, Hawaii.
19. Banwell. H. (1964). *The placing and management of contracts for building and civil engineering work*. A report of the committee under the chairpersonship of Sir Harold Banwell, HMSO, London.

20. Barker, K. (2004). *Review of housing supply. Delivering stability: Securing our housing needs.* A report of the committee under the chairpersonship of Kate Barker, HMSO, London.

21. Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data,* vol. 3. Wiley, Chichester, England.

22. Carmichael, D. G. (1986). "Erlang loading models in earthmoving." *Civil Engineering Systems,* 3(3), 118-124.

23. Carr, R. I. (1979). "Simulation of construction project duration." *J. Constr. Div., ASCE,* 105 (2), 117-128.

24. Chang, D. (1987). *RESQUE.* PhD Thesis, Univ. of Michigan, Ann Arbor, Mich, U.S.A.

25. Cheng, T-M., and Feng, C-W. (2003). "An effective simulation mechanism for construction operations." *Automation in Constr.,* 12 (3), 227-244.

26. Cheng, T-M., Feng, C-W., and Chen, Y-L. (2005). "A hybrid mechanism for optimising construction simulation models." *Automation in Constr.* 14 (1), 85-98.

27. Clark, A. (2002). "Light at the end of the tunnel." *The Guardian,* Thursday 3rd October.

28. Clark, A. (2003). "Winsor adds £8bn to rail repair bill." *The Guardian,* Saturday 18th October.

29. Clayton, M. J., Warden, R. B., and Parker, T. W. (2002). "Virtual construction of architecture using 3D CAD and simulation." *Automation in Constr.,* 11 (2), 227-235.

30. Clemmens J. P., and Willenbrock J. H. (1978). "The SCRAPESIM computer simulation." *J. Constr. Div., ASCE,* 104 (4), 419-435.

31. Cleveland, W. S., and Kleiner, B. (1975). "A graphical technique for enhancing scatterplots with moving statistics". *Technometrics,* vol. 17, pp. 447-454.

32. Collins Bartholomew Ltd. (2003). www.multimap.com. (Accessed: 02/06/2005).

33. Constructing Excellence. (2004). *"About constructing excellence."* www.constructingexcellence.org.uk/about. (Accessed: 02/06/2005).

34. Construction Monitor. (2003). *"Lean construction pilot scheme increases productivity by 40%."* Published on behalf of Department for Trade and Industry (DTi) by Building, London, March 2003.

35. C.T.R.L. (2005). *"Channel tunnel rail link."* http://www.ctrl.co.uk. (Accessed: 02/06/2005).

36. Daniel, C., and Wood, F. S. (1971). *Fitting Equations to Data.* Wiley, New York.

37. Dawood, N. (2005). "Overview of construction simulation at University of Teesside." *Unpublished seminar, Lean Research Group, University of California, Berkeley,* U.S.A., 11th April 2005.

38. Dawood, N., and Marasini, R. (2003). "Visualisation of a stockyard layout simulator "SimStock": a case study in precast concrete products industry." *Automation in Constr.,* 12 (2), 113-122.

39. Dawood, N., Spriprasert, E., Mallasi, Z., and Hobbs, B. (2003). "Development of an integrated information resource base for 4D/VR construction process simulation." *Automation in Constr.,* 12 (2), 123-131.

40. Denny, C., and Halls, S. (2002). "Prescott's housing blueprint under attack." *The Guardian,* Monday 12th August.

41. D.f.T. (2004). "Rail review white paper – the future of rail." HMSO, London.

42. Draper, N. R., and Smith, H. (1998). *Applied Regression Analysis*, vol. 3. Wiley, New York.

43. DTi. (2003). "*Review of early estimates of construction output for GDP in 2003.*" HMSO, London.

44. DTi. (2004). "*Construction statistics annual 2004.*" HMSO, London.

45. Dunlop, P. (2000). Unpublished data sets. University of Edinburgh, U.K.

46. Dunlop, P. (2005). "*Investigation, modelling and planning of stochastic concrete placing operations.*" Unpublished PhD Thesis, University of Edinburgh, U.K.

47. Dunlop, P., and Smith, S. D. (2003). "Estimating the key characteristics of the concrete delivery and placement process using linear regression analysis." *Civil Engineering and Environmental Systems*, 20(4), 273-290.

48. Dzeng, R-J., and Tommelein, I. D. (1997). "Boiler erection scheduling using product models and case-based reasoning." *J. Constr. Engrg. and Mgmt., ASCE*, 123(3), 338-347.

49. Egan, J. (1998). "*Rethinking construction.*" A report of the committee under the chairpersonship of Sir John Egan, HMSO, London.

50. Emerson, H. (1962). "*Survey of problems before the construction industries.*" A report prepared for the Ministry of Works by Sir Harold Emerson, HMSO, London.

51. Faltings, B., Hua, K. F., Schmitt, G., and Shih, S. G. (1991). "Case-based representation of architectural design knowledge." *Proc. DARPA Case-Based Reasoning Workshop*, 307-316.

52. Fellows, R. F., and Liu, A. (2003). *Research methods for construction, 2nd edition.* Blackwell Science, UK.

53. Feng, C-W., Cheng, T-M., and Wu, H-T. (2004). "Optimizing the schedule of dispatching RMC trucks through genetic algorithms." *Automation in Constr.*, 13 (3), 327-340.

54. Flemming, U. (1994). "*Case-based design in the SEED system Knowledge-Based Computer-Aided Architectural Design.*" G. Carrara, and Y.E., Kalay (editors), Elsevier, Amsterdam, The Netherlands, 69-91.

55. Flood, I., and Kartam, N. (1994). "Neural networks in Civil Engineering, I: Principles and Understanding." *J. Comp. in Civ. Engrg., ASCE,* 8, 131-148.

56. Fox, J. (1984). *Linear Statistical Models and Related Methods.* Wiley, New York, U.S.A.

57. Forbes, D.R, McGurnaghan, H., Graham, D., and Smith S.D. (2004). "Concrete placing productivity using a novel neural network design." *ARCOM 20th annual conference*, Heriot-Watt University, Sept 1-3, 1053-1062.

58. Forbes, D. R. (2004). *Private communications between D. R. Forbes and D. Graham.* August 2004.

59. Gagarine, N., Flood, I., and Albrecht, P. (1994). "Computing truck attributes with artificial neural networks." *J. Comp. in Civ. Engrg., ASCE,* 8(2), 179-200.

60. Graham, D., and Smith, S. D. (2005). "A comparison of fuzzy and crisp CBR models of construction processes." *Comp.-Aided Civ. And Infrastruct. Engrg.*, under review.

61. Graham, D., Smith, S. D., and Crapper, M. (2004). "Improving concrete placement simulation with a case-based reasoning input." *Civ. Engrg. and Environ. Systs.*, 21 (2), 137-150.

62. Graham, L. D., Smith, S. D., and Dunlop, P. (2005). "Lognormal distribution provides an optimum representation of the concrete delivery and placement process." *J. Constr. Engrg. and Mgmt., ASCE*, 131 (2), 230-238.

63. Green, S. D. (1999). "The dark side of lean construction." *Proc., 7th Annual Conf. of the Int. Group for Lean Construction*, Brighton, U.K.

64. Hajjar, D., and AbouRizk, S. M. (2002). "Unified modeling methodology for construction simulation." *J. Constr. Engrg. and Mgmt., ASCE*, 128 (2), 174-185.

65. Halpin, D. W. (1973). *An investigation of the use of simulation networks for modeling construction operations.* PhD Thesis, Univ. of Illinois, Urbana-Champaign, Illinois, U.S.A.

66. Halpin, D. W. (1977). "CYCLONE – methods for modeling job site processes." *J. Constr. Div., ASCE*, 103 (3), 489-499.

67. Halpin, D. W. (1990). *MicroCYCLONE User's Manual.* Civil Engineering Dept., Purdue Univ., W. Lafeyette, Indiana, U.S.A.

68. Halpin, D. W. (1993). "Process-based research to meet the international challenge." *J. Constr. Engrg. and Mgmt., ASCE*, 119 (3), 415-425.

69. Halpin, D. W., and Woodhead, R. W. (1976). *Design of construction and process operations.* Wiley, New York, U.S.A.

70. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics- the Approach based on Influence Functions.* Wiley, New York, U.S.A.

71. Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining.* Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.

72. Harriss, C. (1998). "Why research without theory is not research: a reply to Seymour, Crooke and Rooke." *Construction Management and Economics*, 16(1), 113-116.

73. Hawkins, D. M. (1980). *Identification of Outliers.* Chapman and Hall, U.K.

74. Hegazy, T., and Kassab, M. (2003). "Resource optimization using combined simulation and genetic algorithms." *J. Constr. Engrg. and Mgmt., ASCE*, 129 (6), 698-705.

75. Holti, R., Nicolini, D., and Smalley, M. (2000). *The handbook of supply chain management.* CIRIA, London, U.K.

76. Hong, Z., Shi, J. J., and Tam, C. M. (2002). "Visual modeling and simulation for construction operations." *Automation in Constr.*, 11 (1), 47-57.

77. Huang, R. Y., and Halpin, D. W. (1994). "Visual construction operations simulation – the DISCO approach." *J. MicroComp. In Civ. Engrg.*, 9, 175-184.

78. Huang, R. Y., Grigonadis, A. M., and Halpin, D. W. (1994). "Simulation of cable-stayed bridges using DISCO." *Proc. Winter Sim. Conf., IEEE*, Piscataway, N.J., U.S.A.

79. Huang, R. Y., Chen, J-J., and Sun, K-S. (2004). "Planning gang formwork operations for building construction using simulation." *Automation in Constr.*, 13 (6), 765-779.

80. International Bank for Reconstruction and Development (IBRD). (1973). *A framework for the promotion of construction industries in developing countries.* Bank staff working paper #168.

81. Ioannou, P. G. (1984). *The economic value of geological exploration as a risk reduction strategy in underground construction*. PhD Thesis, Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.

82. Ioannou, P. G. (1989). "Dynamic probabilistic decision processes." *J. Constr. Engrg. and Mgmt., ASCE*, 115 (2), 237-257.

83. Isidore, L. J., and Back, W. E. (2002). "Multiple simulation analysis for probabilistic cost and schedule integration." *J. Constr. Engrg. and Mgmt., ASCE*, 128 (3), 211-219.

84. Jackson-Robbins, A. (1998). *Selecting contractors by value*. CIRIA, London, U.K.

85. Jeng, B. C., and Liang, T-P. (1995). "Fuzzy indexing and retrieval in case-based systems." *Expert Systems with Applications*, 8(1), 135-142.

86. Kalk, A. (1980). *INSIGHT: Interactive simulation of construction operations using graphical techniques.* Technical Report No. 238, Dept. Civil Engineering, Stanford Univ., U.S.A.

87. Karshenas, S., and Xin, F. (1992). "Application of neural networks in earthmoving equipment production estimating" *Proceedings of the 8th Conference in Computing in Civil Engineering and Geographic Information Systems, ASCE*, New York, NY, 841-847.

88. Kartam, N. (1996). "Neural network spreadsheet integration for earth-moving operations." *Microcomputers in Civ. Engrg.*, 11(4), 283-288.

89. Kim, K. J., and Gibson Jr., G. E. (2003). "Interactive simulation modeling for heavy construction operations." *Automation in Constr.*, 12 (1), 97-109.

90. Kolodner J. (1993). *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, California, U.S.A.

91. Koskela, L. (1992). *Application of the new production theory to construction*. Technical Report #72, Centre for Integrated Facilities Engineering, Stanford Univ., California.

92. Koskela, L., and Howell, G. (2002). "The underlying theory of project management is obsolete." *Proc. PMI. Research Conf. 2002. Slevin, D., Cleland, D., and Pinto, J. (eds.), Project Management Institute (PMI)*, 293-302.

93. Kottegoda, N. T., and Rosso, R. (1998). *Statistics, probability and reliability for civil and environmental engineers*. McGraw-Hill, Singapore.

94. Kriegsman, M., and Barletta, R. (1993). "Building a case-based help desk application." *IEEE Intell. Syst.*, 8(6), 18-26.

95. Kumar, H., and Krishnamoorthy, C. (1995). "A framework for case-based reasoning in engineering design." *Artif. Intell. For Engrg. Design, Analysis, and Manufact.*, 9, 161-182.

96. Latham, M. (1994). *Constructing the team*. A report of the committee under the chairpersonship of Sir Michael Latham, HMSO, London.

97. Law, A. M., and Kelton, W. D. (2000). *Simulation Modeling and Analysis, vol. 3*. McGraw-Hill, Inc, New York, N.Y., U.S.A.

98. Leake, D. (1996). *Case-Based Reasoning: Experiences, Lessons, and Future Directions.* D. Leake (editor), AAAI Press/MIT Press, Menlo Park, Calif., U.S.A.

99. Leu, S-S., and Hung, T-H. (2002). "A genetic algorithm-based optimal resource-constrained scheduling simulation model." *Constr. Mgmt. and Econ.*, 20 (2), 131-141.

100. Lindley, D. V., and Scott, W. F. (1984). *New Cambridge Statistical Tables, vol. 2*. Cambridge University Press, Cambridge, U.K.

101. Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* Wiley, New York, U.S.A.

102. Liu, L. Y. (1991). *COOPS: Construction object-orientated simulation system.* PhD Thesis, Univ. of Michigan, Michigan, U.S.A.

103. Liu L. Y., and Ioannou P. G., Graphical object-orientated discrete-event simulation system *Proc., 1992 Winter Simulation Conf.*, IEEE, Piscataway, N.J., U.S.A., 1992.

104. Lluch, J., and Halpin, D. W. (1982). "Construction operations and microcomputers." *J. Constr. Div., ASCE*, 108 (1), 129-145.

105. Lu, M. (2002). "Enhancing project evaluation and review technique simulation through artificial neural network-based input modeling." *J. Constr. Engrg. and Mgmt., ASCE*, 128 (5), 438-445.

106. Lu, M. (2003). "Simplified discrete-event simulation approach for construction simulation." *J. Constr. Engrg. and Mgmt., ASCE*, 129 (5), 537-546.

107. Lu, M., Anson, M., Tang, S. L., and Ying, Y. C. (2003). "HKCONSIM: A practical simulation solution to plan concrete plant operations in Hong Kong." *J. Constr. Engrg. and Mgmt., ASCE*, 129 (5), 547-554.

108. Maio, C., Schexnayder, C., Knutson, K., and Weber, S. (2000). "Probability distribution functions for construction simulation." *J. Constr. Engrg. and Mgmt., ASCE*, 126 (4), 285-292.

109. Malkawi, A. (2004). "Developments in environmental simulation." *Automation in Constr.*, 13 (4), 437-445.

110. Manly, B. F. J. (1994). *Multivariate Statistical Methods, vol. 2.* Chapman and Hall, U.K.

111. Martinez, J. C. (1996). *STROBOSCOPE: State and resource based simulation of construction processes.* PhD Thesis, Univ. of Michigan, Michigan, U.S.A.

112. Martinez, J. C. and Ioannou, P. G. (1999). "General-purpose system for effective construction simulation." *J. Constr. Engrg. and Mgmt., ASCE*, 125 (4), 265-276.

113. Marzouk, M., and Moselhi, O. (2002). "Simulation optimization for earthmoving operations using genetic algorithms." *Constr. Mgmt. and Econ.*, 20 (6), 535-543.

114. Mendenhall, W., and Sincich, T. (1995). *Statistics for Engineering and the Sciences, vol. 4.* Prentice-Hall, New Jersey, U.S.A.

115. Moavenzadeh, F., and Markow, M. (1976). "Simulation model for travel costs." *J. Constr. Div., ASCE*, 102 (1), 417-432.

116. Montgomery, D. C., and Runger, G. C. (1994). *Applied statistics and probability for engineers.* Wiley, New York, U.S.A.

117. Morcous, G., Rivard, H., and Hanna, A. M. (2002). "Case-based reasoning system for modeling infrastructure deterioration." *J. Comp. in Civil. Engrg., ASCE*, 16 (2), 104-114.

118. Morua, E. (1986). *Resource strategies for dynamic construction project management.* PhD Thesis, Dept. Civ. Engrg., Univ. of Michigan, Ann Arbor, Mich., U.S.A.

119. Myles, A. J. (1997). *Methods for addressing some practical issues in MLP regression and their application to modelling curl in papermaking.* PhD Thesis, Electrical Engrg. Dept., Univ. of Edinburgh, U. K.

120. Naji, K. K. (2000). "Simulating equipment –based construction operations using virtual environments." *Proc. 17th Ann. Symposium on Automation and Robotics in Constr.*

121. Naji, K. K., and Najafi, F. (1996). "The future of virtual reality in simulation of construction operations." *Proc. ASCE Ann. Conf.*

122. Nassar, K., Thabet, W., and Beliveau, Y. (2003). "Simulation of asphalt paving operations under lane closure conditions." *Automation in Constr.*, 12 (5), 527-541.

123. Naylor, T. H. (1971). *Computer simulation experiments with models of economic systems.* Wiley, New York, U.S.A.

124. Neville, A. M., and Kennedy, J. B. (1964). *Basic statistical methods for engineers and scientists.* International Textbook Co., Scranton, P.A., U.S.A.

125. National Ready Mixed Concrete Association (NRMCA). (2001). *Industry fact sheet.* http://nrmca.org. (Accessed: 04/09/2005).

126. Odeh, A. M. (1992). Construction integrated planning and simulation model. PhD Thesis, Dept. of Civ. Engrg., Univ. of Michigan, Mich., U.S.A.

127. Ohno, T. (1987). *Toyota production system.* Productivity Press, New York., U.S.A.

128. Opdenbosch, A., and Baker, N. (1994). *Simulation of construction operations in virtual interactive environments.* PhD Thesis, Georgia Inst. Tech., Georgia, U.S.A.

129. Oxford English Dictionary. (1996). *The Oxford English Dictionary.* Oxford Univ. Press, Oxford, U.K.

130. Ozel, F., and Kohler, N. (2004). Data modeling issues in simulating the dynamic processes in life cycle analysis of buildings. *Automation in Constr.*, 13 (2), 167-174.

131. Palisade Corporation. (2003). *Guide to Bestfit.* http://www.palisade-europe.com. (Accessed: 02/06/2005).

132. Paulson, B. (1978). "Interactive graphics for simulating construction operations." *J. Constr. Div., ASCE*, 104 (3), 69-76.

133. Paulson, B., Chan, W., and Koo, C. C. (1987). "Construction operations simulation by microcomputer." *J. Constr. Engrg. and Mgmt., ASCE*, 113 (2), 302-314.

134. Penrose, L. W. (1953). "Distance, size and shape." *Annals of Eugenics*, 18, 337-343.

135. Pritsker, A. A. B. (1998). "Life and death decisions: Organ transportation policy and analysis." OR/MS Today, August 1998.

136. Pritsker, A. A. B., O'Reilly, J. J., and LaVal, D. K. (1997). *Simulation with Visual SLAM and AWESIM.* Systems Publishing Corp., New York, U.S.A.

137. Pullen, A., Newman, J. B., and Chana, P. (2000). *Process efficient concreting: improved speed and quality.* CRC Ltd, Berkshire, U.K.

138. Raferty, J. (1998). "From Ptolemy to Heisenberg: quantitative models and reality." *Construction Management and Economics*, 16(3), 295-302.

139. Raferty, J., McGeorge, D., and Walters, M. (1997). "Breaking up methodological monopolies: a multi-paradigm approach to construction management research." *Construction Management and Economics*, 15(3), 291-297.

140. Reich, Y., and Fenves, S. J. (1995). "System that learns to design cable-stayed bridges." *J. Infrastruct. Syst.*, 121 (7), 1090-1100.

141. Reichard, G., and Papamichael, K. (2005). "Decision-making through performance simulation and code compliance from the early scheme phases of building design." *Automation in Constr.*, 14 (2), 173-180.

142. Ren, A., Wen, Y., Chen, C., and Shi, J. J. (2004). "Modeling of irregular structures for the construction simulations in virtual reality environments based on the web." *Automation in Constr.*, 13 (5), 639-649.

143. Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, U.K.

144. Rivard, H., Fenves, S. J., and Gomez, N. (1988). "Case-based reasoning for conceptual building design." *1st Int. Conf. New Information Tech. Decision-Making in Civ. Engrg.*, Montreal, Canada, 355-366.

145. Roddis, W. M. K., and Bocox, J. (1997). "Case-based approach for steel fabrication errors." *J. Comp. in Civ. Engrg., ASCE*, 11 (2), 84-91.

146. Rooke, J., Seymour, D., and Crook, D. (1997). "Preserving methodological consistency: a reply to Raferty, McGeorge and Walters." *Construction Management and Economics*, 15(5), 491-494.

147. Rowings Jr., J. E., and Sonmez, R. (1996). "Labor productivity modeling with neural networks." *Transactions of AACE International*, PRD11.

148. Rowings Jr., J. E., and Sonmez, R. (1998). "Construction labor productivity modeling with neural networks." *J. Constr. Engrg. and Mgmt., ASCE*, 124 (6), 498-504.

149. Runeson, G. (1997). "The role of theory in construction management research: a comment." *Construction Management and Economics*, 15(3), 299-302.

150. Santos-Pereira, C. M., and Pires, A. N. (2002). "Detection of outliers in multivariate data: a method based on clustering and robust estimators", *In Hardle, W., and Ronz, B. (Eds.), Proc. in Computational Statistics: 15th Symposium*, Berlin, Germany, Physica-Verlag, Heidleberg, 291-296.

151. Seber, G. A. F. (1984). *Multivariate Observations*. Wiley, New York, U.S.A.

152. Sexty, C., and Sefton, T. (2004). "Hitting the right target." *The Guardian*, Friday 21st May.

153. Seymour, D., and Rooke, J. (1995). "The culture of the industry and the culture of research." *Construction Management and Economics*, 13, 511-523.

154. Seymour, D., Crook, D., and Rooke, J. (1997). "The role of theory in construction management: a call for debate." *Construction Management and Economics*, 15(1), 117-119.

155. Seymour, D., Crook, D., and Rooke, J. (1998). "The role of theory in construction management: a reply to Runeson." *Construction Management and Economics*, 16(1), 109-112.

156. Shanks, R., and Abelson, R. (1977). *Scripts, plans, goals and understanding*. Lawrence Erlbaum Assoc., Hillsdale, N.J., U.S.A.

157. Shi, J. J. (1999). "Activity-based construction (ABC) modeling and simulation method." *J. Constr. Engrg. and Mgmt., ASCE*, 125 (5), 354-360.

158. Shi, J. J. (2002). "Three methods for verifying and validating the simulation of a construction operation." *Constr. Mgmt. and Econ.*, 20 (6), 483-491.

159. Shih S. G. (1991). *Case-based representation and adaptation in design*. In: CAAD futures '91, G.N., Schmitt (editor), Braunshweig: Vieweg Verlag, Germany.

160.Simon, E. (1944). *The placing and management of building contracts. A* report of the Central Council for Works and Buildings under the chairpersonship of Lord Earnest Simon, HMSO, London, U.K.

161.Smith, S. D. (1992-1994). Unpublished data sets. Univ. of Edinburgh, U.K.

162.Smith, S. D. (1995). *Productivity estimation of earthmoving operations using discrete-event simulation.* PhD Thesis, Univ. of Edinburgh, U.K.

163.Smith, S. D. (1998). "Concrete placing analysis using discrete-event simulation." *Proc. Inst. Civ. Engrg., Structures and Buildings*, 128 (4), 351-358.

164.Smith, S. D. (1999). "Modeling and experimentation of the concrete supply and delivery process." *Civ. Engrg. and Environ. Syst.*, 16 (2), 93-114.

165.Smith, S. D., Osbourne, J. R., and Forde, M. C. (1995). "The use of a discrete-event simulation model with Erlang probability distributions in the estimation of earthmoving production." *Civ. Engrg. and Environ. Syst.*, 13 (1), 25-44.

166.Sprent, P. (1993). in *Applied Nonparametric Statistical Methods*, vol. 2, 277-293. Chapman and Hall, U. K.

167.Stephens, M. A. (1974). "EDF statistics for goodness of fit and some comparisons." *J. Am. Stat. Assoc.*, 69, 730-737.

168.Tah, J. H. M., Carr, V., and Howes, R. (1999). "Information modeling for case-based construction planning of highway bridge projects." *Adv. Engrg. Software*, 30, 495-509.

169.Teicholz, P. (1963). *A simulation approach to the selection of construction equipment.* Tech. Report 26, The Construction Inst., Stanford Univ, U.S.A.

170.The Pyramid Center. (2005). The Pyramid Center. http://www.tapyramid.com. (Accessed: 02/06/2005).

171.Thomas, H. R., Horman, M. J., de Souza, U. E. L., and Zavrski, I. (2002). "Reducing variability to improve performance as a lean construction principle." *J. Constr. Engrg. and Mgmt., ASCE*, 128 (2), 144-154.

172.Tommelein, I. D. (1998). "Pull-driven scheduling for pipe-spool installation: Simulation of lean construction technique." *J. Constr. Engrg. and Mgmt., ASCE*, 124 (4), 279-288.

173.Touran, A., and Asai, T. (1987). Simulation of tuneling operations." *J. Constr. Engrg. and Mgmt., ASCE*, 113 (4), 554-568.

174.Tukey, J. (1977). *Exploratory Data Analysis.* Addison-Wesley, Mass., U.S.A.

175.Wakefield, R. R. (1998). "Application of extended stochastic Petri nets to simulation and modeling of construction systems." *Civ. Engrg. and Environ. Syst.*, 15 (1), 1-22.

176.Wales, R. J., and Abourizk, S. M. (1996). "Integrated simulation model for construction." *Simulation Practice and Theory*, 3 (6), 401-420.

177.Watson I. (1997). *Applying Case-Based Reasoning: Techniques for Enterprise Systems.* Morgan Kaufmann, San Mateo, Calif., U.S.A.

178.Weaver, M. (2002). "Prescott earmarks brownfield sites for new homes." *The Guardian*, Wednesday 24th July.

179.Wing, C-K., Raferty, J., and Walker, A. (1998). "The baby and the bath water: research methods in construction management." *Construction Management and Economics*, 16(1), 99-104.

180. Womack, J. P., and Jones, D. T. (1996). *Lean thinking: Banish waste and create wealth in your corporation*. Simon and Shuster, New York, U.S.A.

181. Womack, J. P., Jones, D. T., and Roos, D. (1990). *The machine that changed the World*. Maxwell MacMillan International, Oxford, U.K.

182. Woods, D. G., and Harris, F. C. (1980). "Truck allocation for concrete distribution." *J. Constr. Div., ASCE*, 106 (2), 131-139.

183. World Bank. (1984). *The construction industry, issues and strategies in developing countries*. World Bank, Washington D.C., U.S.A.

184. Yang, J., Edwards, D. J., and Love, P. E. D. (2003). "A computational intelligent fuzzy model approach for excavator cycle time simulation." *Automation in Constr.*, 12 (6), 725-735.

185. Zhang, H., Shi, J. J., and Tam, C. M. (2002a). "Iconic animation for activity-based construction simulation." *J. Comp. in Civ. Engrg., ASCE*, 16 (3), 157-164.

186. Zhang, H., Shi, J. J., and Tam, C. M. (2002b). "Application of simulation related techniques to construction operations." *Engrg. Constr. And Architect. Mgmt.*, 9 (6), 433-445.

187. Zhang, H., Tam, C. M., and Shi, J. J. (2003). "Application of fuzzy logic to simulation for construction operations." *J. Comp. in Civ. Engrg., ASCE*, 17 (1), 38-45.

188. Zhang, H., Li, H., and Tam, C. M. (2004). "Fuzzy discrete-event simulation for modeling uncertain activity duration." *Engrg. Constr. And Architect. Mgmt.*, 11 (6), 426-437.

Stochastic modelling and analysis of construction processes

Appendices

Darren Graham

PhD Thesis – University of Edinburgh

List of Appendices

APPENDIX A: MORE ABOUT LEAN CONSTRUCTION 225
APPENDIX B: APPLICATION OF THE SIMPLE OUTLIER IDENTIFICATION METHOD 235
APPENDIX C: THE RMC SUPPLY PROCESS VARIABLES 241
APPENDIX D: SOURCE CODE OF COMPUTER MODELS 247
APPENDIX E: CALCULATION OF THE PROCESS PERFORMANCE MEASURES 273
APPENDIX F: FIT OF PROBABILITY DISTRIBUTIONS TO THE TIMES OF INTEREST 277
APPENDIX G: VALIDATION OF *ExcelSim* 281
APPENDIX H: MATHEMATICAL NOTATION 283
APPENDIX I: PUBLICATIONS 285

APPENDIX A

MORE ABOUT LEAN CONSTRUCTION

A.1 Introduction

This purpose of this appendix is to provide a definition of lean construction and to discuss the reaction of the UK construction industry to this new managerial concept.

A.2 What is Lean construction?

Lean construction is the application of lean production management methods in construction projects, and was pioneered by Lauri Koskela of VTT Technology, Finland in the late 1980s and early 1990s (Ballard, 2000). Lean production management was first developed in the Japanese automobile industry, at Toyota, and has since caused a revolution in manufacturing design, supply and assembly (Womack et al., 1990; Womack and Jones, 1996). The overall objective of lean production management (and lean construction) is to maximise the *value* of a product, whilst minimising the waste involved in the production of that product. This is achieved through concurrently optimising both a product's design and the processes involved in producing a product, to result in an improvement in the total project performance.

The meaning of *value* in the pre-lean, automobile and general manufacturing industries was to provide a client with a product at the lowest possible cost, with a seeming disregard to productivity, time, quality, and client specification. Contracts for work were (and still often are in construction) won by providing a tender bid for the lowest price, especially in government projects, which are prevalent in the present UK construction industry. Lean production management challenges this traditional concept of *value*. *Value* in lean terms was defined by Womack et al. (1990) as: 'providing a client with the product they have ordered in the shortest possible time, at the lowest possible cost and with the highest possible quality'. Thus, contracts are no longer won by the lowest bidder, but by the individual/group who offer the most value.

Traditional production theories, such as mass or craft production, consider production to consist of conversions (activities) of a raw material into something closer to the final product further along the production process line. Traditionally, every activity is considered necessary to the completion of the product. For instance, earth must be removed from almost every project in the formation of foundations, etc. On a large-scale excavation of earth, a back-actor (loader) is often used to scoop up the earth and several dump trucks (haulers) are used to transport the material to a specified area for dumping the material. In this case the final

product is a hole in the ground and the following are examples of the conversions (from a traditional viewpoint) that are necessary to form this product (shown in figure A.1):

- To begin, a hauler is position adjacent to the loader (conversion 1).
- Next, the loader fills up the hauler (conversion 2).
- The hauler then transports this load to the area designated to hold the material (conversion 3).
- The hauler dumps the material (conversion 4).
- Subsequently the hauler returns to the loading area (conversion 5).
- As there are many haulers working within the system, the likelihood is that a queue of haulers may form at the loader – as only one hauler is usually loaded at a given time. This queuing is undoubtedly a source of frustration for a contractor, but is seen as a consequence of the process of earthmoving (or as conversion 6 of the process). Only when an especially large amount of time with haulers remaining static is experienced an attempt may be made to reduce the queuing of trucks.
- Finally, to complete the process a hauler is again positioned at the loader (conversion 1).

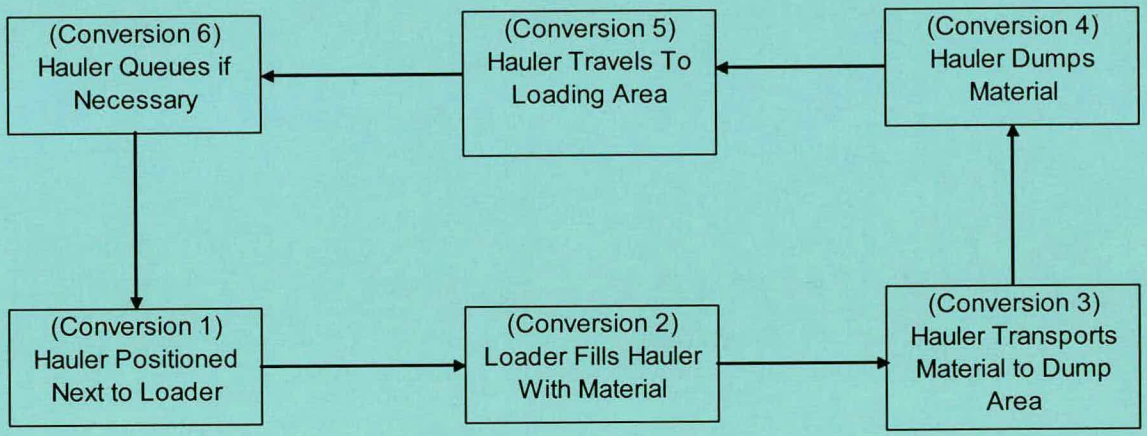


Figure A.1: Traditionally-viewed process of earthmoving

Lean production management takes a very different view of the process of forming a product: production is considered to consist of physical conversions of materials (activities) and flows (of information, resources, etc.) between conversions (Ballard and Howell, 1994) and identifies both value-adding and non value-adding (wasteful) activities and flows in the process. An

ambition of lean production management is to remove (or at least reduce) the waste present in the production process (Ohno, 1987).

The above concepts are illustrated again using the earthmoving process as an example (see figure A.2):

- The first step is to position a hauler next to the loader (flow 1) – this is a value-adding flow because the material is not physically being brought closer to the final product. Rather, the movement of the hauler into position next to the loader is essential to the completion of the process, making this flow process value-adding.
- Next, the loader fills the hauler with material (conversion 1; value-adding).
- When full, the hauler travels to an area designated for holding the earth (flow 2; value-adding).
- The hauler dumps its load in an appropriate area (conversion 2; value-adding).
- Then the hauler returns to the loader (flow 3; value-adding).
- As explained above, it is probable that a queue of haulers may form at the loader (flow 4) – as only one hauler is likely to be loaded at a time. This queuing is a non value-adding flow. In lean production management (and lean construction), the reduction/minimisation of such a non value-adding flow or activity would be essential and actively pursued, provided that this sub-optimisation of the process of taking concrete deliveries did not adversely affect the overall value of the construction project.
- Finally, flow 1 is undertaken by the hauler to complete the earthmoving cycle.

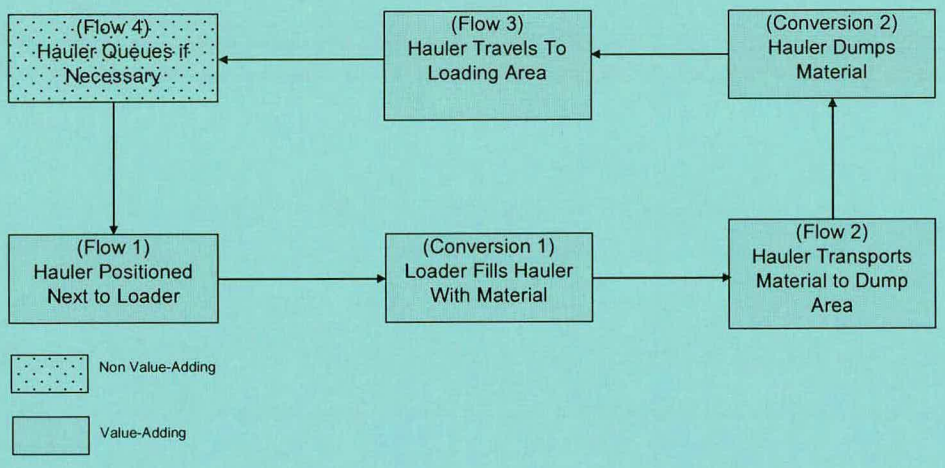


Figure A.2: Lean production management view of the earthmoving process

How can lean production management be implemented practically in construction projects? Ten practical procedures have been suggested that can be used by a contractor to change from a more traditional approach of managing processes to a lean management style (Womack et al., 1990; Womack and Jones, 1996). They have been listed below along with a succinct description of each one:

1. **Increase process transparency**. The activities involved in any process must be made (at least more) transparent, allowing for an easy identification of the parts of a process which are under-performing. This makes individuals who have a responsibility for a specific activity in a process more accountable and (hopefully) will encourage a high level of performance from those individuals.
2. **Introduce continuous process improvement**. Continuous process improvement involves: the employees working in the process to seek out the sources of any problems encountered and prevent any recurrence of the same problem, not just fix a problem and accept that it occurred; and management and employees working together to optimise work practices. For this method to work an attitude change among the employees and management working within the process is required. This coupled with the iterative nature of continuous process improvement can mean it takes decades to optimise the performance of a particular process, as witnessed by the developers of lean production, Toyota. They have been using continuous process improvement techniques since the mid 1940s and this has resulted in a repair area of 4% of the overall assembly space (Womack et al., 1990). When compared with a North American car manufacturer (using mass production management) whose repair area is 13% (Womack et al., 1990) the amount of re-work lean production saves is clear. Getting the product correct the first time is even more imperative in construction than the automobile industry - as an entire construction project could be stalled due a small area of re-work.
3. **Focus process controls on the totality of a process.** Controls should only be placed upon a process if they result in a better 'value' process. Thus, no activity should be optimised, unless this has the effect of producing an increase in the value of a process.
4. **Establish permanent references in the process (benchmarking).** Processes usually have the same set of activities that are required to be performed. Thus it is possible to define a series of generic, permanent reference points in a process and make a

measurement of the process performance (e.g. productivity) at each of them. This allows the performance of an activity to be compared directly with the performance recorded from the same activity in an earlier project, providing a measure of the current activity performance. Action can then be taken to improve the activity if required.

5. **Increase the efficiency of value-adding activities**. In a lean construction project, value-adding activities are those which directly contribute to the completion of a process (e.g. discharging a batch of RMC into a concrete pump). Thus, this type of activity should be performed with the utmost efficiency to maximise the total process value.
6. **Reduce the share of non value-adding activities**. Non value-adding activities (e.g. plant sitting idle on site) are seen as wasteful in lean construction. As they are not adding anything to the total process value, they must be removed (or at least minimised).
7. **Reduce the variability in a process**. This is a key procedure in implementing lean principles as a highly variable process is difficult to plan and control. If the variability in a process can be minimised then the employees and management know (to a high degree) what needs to be done and when. Thus, managers and employees with limited amounts of time can focus more readily upon implementing the remaining implementation procedures of lean production.
8. **Reduce process cycle times**. The time taken to complete a process or the process cycle time is inherently linked with the overall duration of a project. Thus, to reduce the time taken to complete an individual process should have the effect of reducing the duration of a whole project, which in theory provides better value to the client.
9. **Increase the process value through a systematic consideration of a client's requirements**. It is essential to identify exactly who the client's are in a process. Intuitively, one may assume the client of a construction project is the person/entity who will 'receive the keys' on completion of the project. However, if each process that forms a complete project was divided into the activities which constitute a particular process in the systematic order in which these activities must be carried out, many more clients are revealed within an individual process. For instance, part of the earthmoving process can be broken down into:

- A loader (client) and loader operator require that a sufficient number of haulers are present in the process to allow the loader to work constantly. In this instance the need can be fulfilled by a planner.
- A hauler (client) is reliant on the actions of the loader. Clearly, if a hauler is made to remain idle by the loader then the needs of the client, which are to be loaded with earth, are not being met.
- A filled hauler (client) then needs the logistics to be in place to allow unimpeded travel to the dumping area.

The requirements of the client in each of these activities will then be met provided that action increases the value of the total process.

10. **Simplify through the minimisation of the number of process activities.** In any process there are perceived to be a number of activities which must be undertaken to complete the total process. For example:
    - A material being manufactured at a remote location, when ready, must be loaded onto a vehicle and transported to a site.
    - On site it is likely that the material will be stored until it is ready to be used.
    - When the material is required, it will need to be transported again to the appropriate area.
    - Finally, the material can be used in a particular construction process.

    In the process of using this pre-made material 4 activities were required to get the material to the 'right place at the right time'. This number could be reduced to 2 by transporting the material just before it is required and delivering it directly to the appropriate area. This 'just-in-time' management style would remove the need for both 'double-handling' the material and for having a storage area at all. Minimising the number of activities contained within a process is difficult as time becomes a more critical factor, and potentially could result in a construction project without the necessary materials if the supply chain were to fail in some manner.

Has the lean construction research community been advocating the implementation of these practical procedures? The answer is yes, but not fully. Yes, because each of the areas above have been the focus of research by members of the International Group for Lean Construction (IGLC) and the Lean Construction Institute (LCI) (e.g. Alarcon, 1993 and 1994; Koskela, 1992; Ballard and Howell, 2003; Tommelein, 1998). As previously mentioned, an element of lean

construction focuses on, managing the flows between conversions. This management takes the form of controlling the workflow - ensuring that it is steady and stable. Thus, minimising the variability in the conversion processes, and the flows between these processes, is the goal. To reduce process and flow variability, some construction researchers have advocated the use of buffers, which are stockpiles of materials, information, etc. that enable a job to be worked on at a steady rate to completion (e.g Ballard, 1994; Tommelein, 1998; Koskela and Howell, 2002). This practice is known as shielding, and the idea is that once a shield has been installed it is possible to focus on improving the processes downstream (closer to the finished product) of the shield. Shielding is in essence not a lean management method - there is waste involved - and thus, it should not be strictly undertaken in a lean project. In reality, the variability in the construction supply chains, because they can be long and complicated, is such that there is no other option than to shield downstream construction (Ballard and Howell, 1994; 1997).

One of the main methods of improving the processes downstream of the shield is to use 'The Last Planner' system developed by Ballard and Howell at the Lean Construction Institute. The Last Planner involves the people who are undertaking the construction activities to measure the amount of work that has been completed over a period of time (usually one week) and use this measure to form a realistic plan for the work to be undertaken in the subsequent week, reportedly leading to improvements in project performance (Ballard and Howell, 1998). The benefit of using buffers and shields in construction is criticised by Thomas et al. (2002) who suggest that while the results produced by Ballard and Howell (1994; 1997) are useful, they are 'hardly conclusive', and 'more analysis is needed to confirm the causal link between reductions in work flow variability and improvements in project performance.

To re-state, studies commissioned by the UK Government and other institutions have proposed lean construction as a method of improving the performance of the construction industry. Accordingly the question begs, has the UK construction industry 'slimmed down' its attitude to management and adopted lean methods in the post-Egan era?

A.3 How has the UK construction industry reacted to lean construction?

Generally, there is scepticism in the industry and academia about the benefits that lean construction can provide (Construction Monitor, 2003; Green, 1999). Despite this, there have been some significant applications of lean construction in the UK. British Airports Authority

(BAA), one of the UK's major clients, has implemented lean construction methods in their construction projects. Their 'show case' project in which lean methods have been implemented is the construction of a new terminal building (Terminal 5) at Heathrow Airport, London (Lean Construction Institute, 2002). Terminal 5 is a massive project which is being undertaken in two phases – with phase I due to be completed in 2008 and phase II due in 2011 – and it is estimated that Terminal 5 will be able to accommodate 30 million passengers per year (BAA, 2005). The project is so large it needed to be separated into 16 major projects, which were then split into over 100 sub-projects (worth between £30 and £150 million each).

Another major project in the UK which is utilising lean construction methods is the construction of the Channel Tunnel Rail Link (CTRL). CTRL is owned by London and Continental Railways Ltd. who have separated the project into two sections (based upon location) and lean methods have been applied to section 2 of the project, which is located between London St. Pancras and Ebbsfleet, Kent. This section of CTRL is worth £3.3 billion, has created 8,000 jobs and is due to be completed in early 2007 (CTRL, 2005).

The UK Government acting through the Department of Trade and Industry (DTi) has implemented a scheme to promote the adoption of lean construction practices in the industry, as a result of the Latham and Egan reports. This scheme is called the 'Construction Lean Improvement Programme' (CLIP) and involves construction contractors working with lean production experts, to re-design the processes involved in a project, using the methods already discussed in this chapter. CLIP is being implemented in stages: Stage one[3] began in January 2002 and involved 7 small-medium contractors. This stage of CLIP has been completed and productivity was increased (from standard working practices) by up to 40% (Construction Monitor, 2003). Stage two began in March 2003, involving more contractors. The results are yet to be published even though the programme was completed over one year ago.

There may be other instances of lean construction being used in construction projects, whether in a pilot scheme or in a full implementation, but at the moment the author is not aware of any.

[3] The contractors involved were: NG Bailey (a mechanical-electrical contractor), Thomas Vale, Denne Construction, Simpson York, Stepnell, Keepmoat and Cruden Construction. The projects varied in type from schools to social housing development, and varied in value from thousands up to £2.6 million.

APPENDIX B

APPLICATION OF THE SIMPLE OUTLIER IDENTIFICATION METHOD

B.1 Introduction

The aim of this example is to assess the effectiveness of using the simple outlier detection method. This assessment shall be made by comparing the results of the simple method with those from the method of Santo-Pereira and Pires (2002) (see 5.3.2) and the known outliers in the data.

The data set has been taken from Rothamsted Experimental Station, one of the oldest agricultural research centres. One of the experiments they carry out at Rothamsted is to monitor the growth of grasses under various fixed treatment regimes. This experiment has been conducted for the past 150 years, providing a substantial data set. Table B.1 presents the yields of grass (in t/ha dry matter) on two totally untreated plots for the 50 years from 1941 until 1990.

Year	Plot 3	Plot 12	Year	Plot 3	Plot 12
1941	0.85	1.26	1966	1.43	2.16
1942	0.26	0.59	1967	1.31	1.48
1943	1.03	1.66	1968	1.52	1.28
1944	0.34	0.65	1969	0.72	1.87
1945	1.14	1.75	1970	1.15	1.51
1946	1.18	0.80	1971	1.51	2.94
1947	1.52	1.67	1972	1.40	1.54
1948	1.12	1.25	1973	1.24	1.27
1949	0.62	0.78	1974	1.18	1.25
1950	0.89	0.76	1975	0.91	0.55
1951	1.00	1.42	1976	1.06	1.22
1952	1.58	1.80	1977	1.20	1.21
1953	1.63	1.84	1978	1.70	1.77
1954	0.99	1.05	1979	1.26	2.27
1955	1.10	1.58	1980	0.85	1.07
1956	0.69	1.11	1981	1.47	1.95
1957	0.60	1.02	1982	2.03	1.91
1958	1.21	1.61	1983	0.99	0.84
1959	0.51	0.62	1984	1.08	1.22
1960	1.56	1.82	1985	1.58	1.65
1961	1.39	1.82	1986	0.78	1.02
1962	1.20	1.28	1987	1.39	1.71
1963	1.43	1.64	1988	1.40	1.38
1964	1.48	2.47	1989	0.60	0.75
1965	2.75	3.45	1990	0.88	0.94

Table B.1: Yields of grass of two untreated plots

Barnett and Lewis (1994) conducted an outlier detection analysis of the grass data, using the Mahalanobis distance measurement, under the assumption of a normally distributed data set. They found that the information from the following years contained suspected outliers: 1965, 1971, 1969, 1982, 1942, 1964, and 1979, which are labelled A to G in figure B.1.

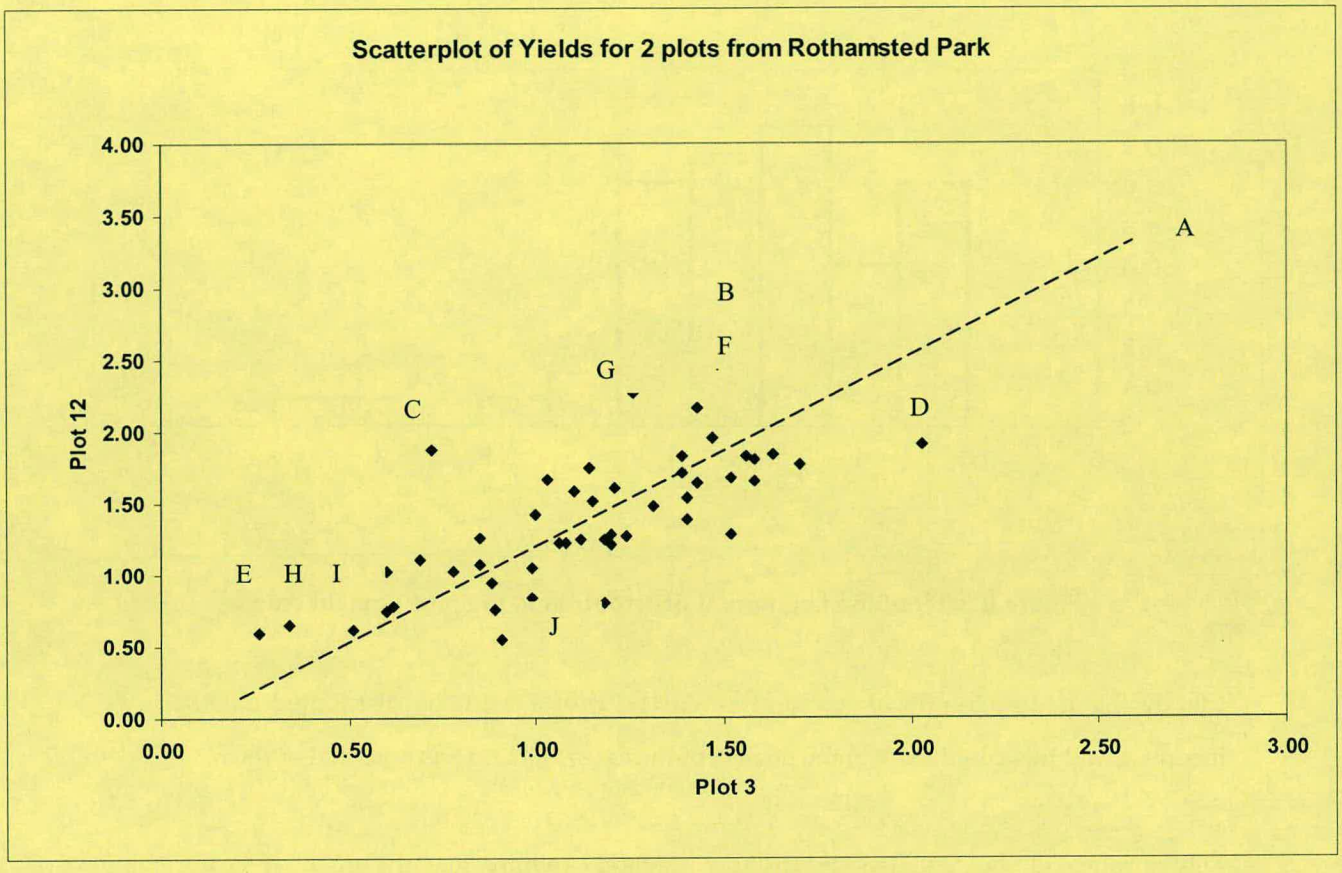


Figure B.1: Scatterplot of the yields from the two untreated grass plots

As the grass data set is bivariate it is possible to scatterplot the variables and examine the plot visually to determine the presence of outliers. In addition to the seven suspected outliers already detected by Barnett and Lewis (1994) (see above), a visual inspection has added further suspected outliers. They are the cases: H, I, and J (see figure B.1) which correspond to the years 1944, 1959 and 1975, respectively. This gives 10 suspected outliers in the grass data set.

Before the simple outlier detection method can be used it is important to determine the probability distributions that underlie the data (see 5.2). This examination was undertaken using the same method as in section 5.2. To summarise the method, a number of theoretical probability distribution were fitted to the data and the distribution which represented the data

best was assessed using the AD and KS tests. The lognormal distribution was found to have the best fits for both plots 3 and 12. Figure B.2 shows the lognormal fit to the data of plot 12.

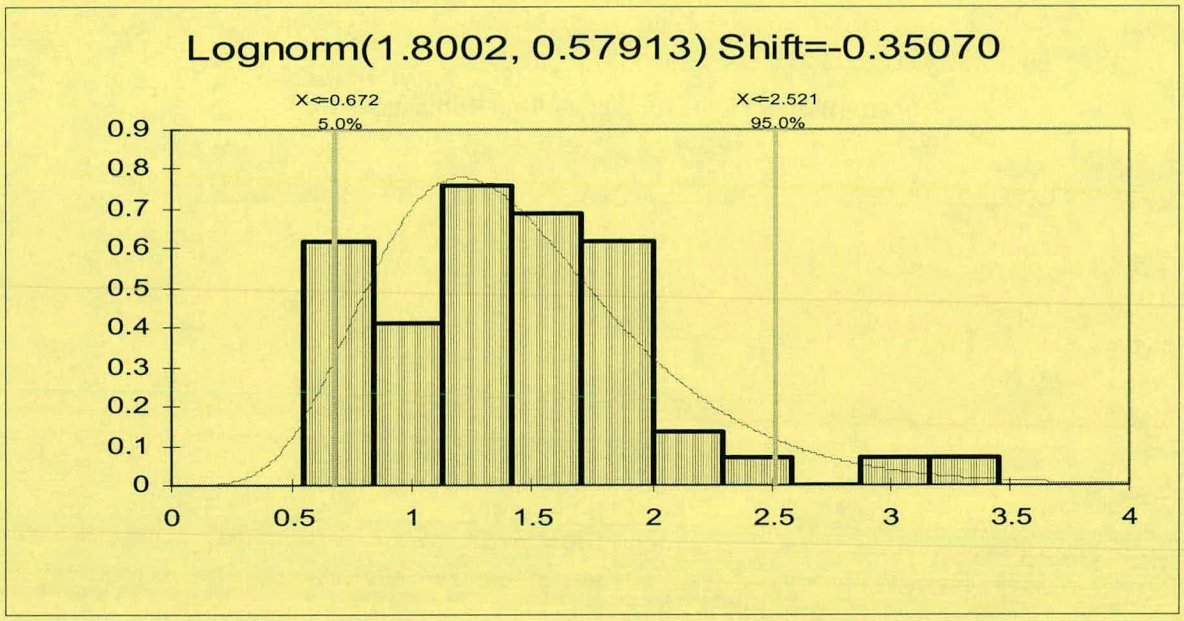


Figure B.2: Fit of the Lognormal distribution to the plot 12 yield data

Clearly the assumption by Barnett and Lewis (1994) of a normally distributed data set was incorrect, and this could explain the neglect of the cases H, I and J as potential outliers.

The simple and the Mahalanobis distance methods require that the data set is normally distributed (see 5.2), which is not the case with the grass data. To make the data more normally distributed a k-means cluster analysis was undertaken (see 5.3.2) using k-values of 2, 3, and 4 - to ensure a good picture of the effect on the data of cluster analysis is presented. The consequences of cluster analysis on the underlying distribution of the data are shown in table B.2. Applying a cluster analysis to the grass data has changed the underlying distributions, and using k = 3 results in the normal distribution becoming the best representative of plot 3. Following cluster analysis the correlations of the variables were examined to allow the Penrose distance measure to be used (see 5.3.2). Table B.3 shows the correlation coefficients of the Pearson, Kendall and Spearman correlation tests and it can be concluded that although the coefficients are not insignificant, the variables are not correlated strongly and the Penrose measure can be used to identify outliers in the data set.

			Number of Clusters											
			2			3			4					
		Cluster Number	1		2	1		2	3		1		2		3		4	
Goodness of Fit Test	Plot 3	Plot 12	Plot 3	Plot 12	Plot 3 Plot 12	Plot 3 Plot 12		Plot 3 Plot 12		Plot 3 Plot 12		Plot 3 Plot 12		Plot 3 Plot 12		Plot 3	Plot 12
A-D	lognormal	lognormal	lognormal	exponential	weibull normal	~	~	normal uniform		normal lognormal		~	~	weibull uniform		normal exponential		lognormal exponential
K-S	lognormal	lognormal	lognormal	exponential	weibull normal	~	~	normal uniform		weibull lognormal		~	~	weibull uniform		normal exponential		lognormal exponential

Table B.2: Distributions that best represent the yields of plots 3 and 12 of the grass data set

Correlation Test		Variable	
Pearson		Plot 3	Plot 12
	Plot 3	1	0.79
	Plot 12	1	0.79
Kendall			
	Plot 3	1	0.61
	Plot 12	0.61	1
Spearman			
	Plot 3	1	0.78
	Plot 12	0.78	1

Table B.3: Correlation test results for the grass data variables

Number of Clusters	Simple Measurement	Mahalanobis Measurement	Suspected Outliers
2	A	A	A
	B		B
	D		C
	E		D
	F		E
	G		F
	H		G
	I		H
	J		I
			J
Total Suspected Outliers	9	1	10
3	A	A	
	B	E	
	D	H	
	E	I	
	F	J	
	G		
	H		
	I		
	J		
Total Suspected Outliers	9	5	
4	A	A	
	B	E	
	D	H	
	E	I	
	F	J	
	G		
	H		
	I		
	J		
Total Suspected Outliers	9	5	

Table B.4: Results of outliers detection analysis

The results of the outlier detection analysis are shown in table B.4. The simple method detects almost all of the known outliers to be in the data set, and this method clearly outperforms the Mahalanobis distance method (at the 1% significance level). Therefore, the conclusion of this study is that the simple method should be used for identifying outliers in the construction data sets.

APPENDIX C

THE RMC SUPPLY PROCESS VARIABLES

C.1 Introduction

As mentioned in 5.3.2, it is unclear whether a linear model would provide adequate productivity estimates of the RMC supply process. The linear model (5.3.2) was overestimating productivity for lower observed productivities and underestimating for higher observed productivities (see figure 5.2). Therefore, it is necessary to undertake a detailed investigation to establish the adequacy of a linear (or nonlinear) method of modelling concreting productivity.

C.2 Is the relationship described better by a nonlinear model than a linear one?

To form an answer to this question the following techniques have been used:

- **Apply a linear model to the data and examine for a nonlinear trend in the fit residuals.** Figure C.1 shows a scatterplot of the fit residuals against the productivity estimated by the linear model, with an 11-point running mean to indicated the trends in the data. This plot gives an indication that a nonlinear relationship exists in the data, as there are a number of local minima, an overall local minima, and the running mean is evidently not a straight line. To provide further evidence of a nonlinear relationship other measures of statistical trends were employed on the data: midmean, and upper and lower semi-midmeans (see 5.3.3 for more information about these measures). Figure C.2 shows a scatterplot of residuals vs. estimates with these statistical measures. There is a suggestion of a nonlinear trend in the data as the measures arrive at a number of local minima and maxima. Although this method has provided an indication of the nature of the process productivity data, it can be oblivious to certain types of nonlinearity. Therefore, another method, which provides more robust evidence of nonlinearity in the data, has been employed.

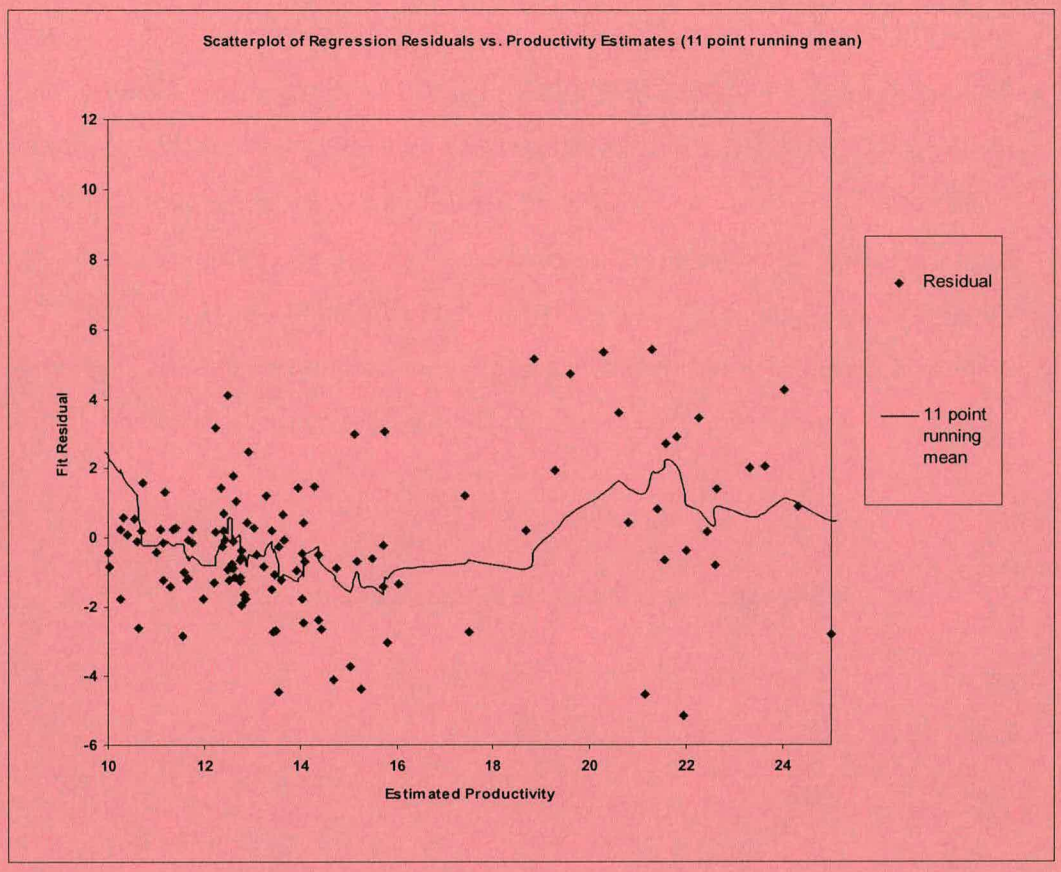


Figure C.1: Scatterplot of the linear model residuals vs. productivity estimates

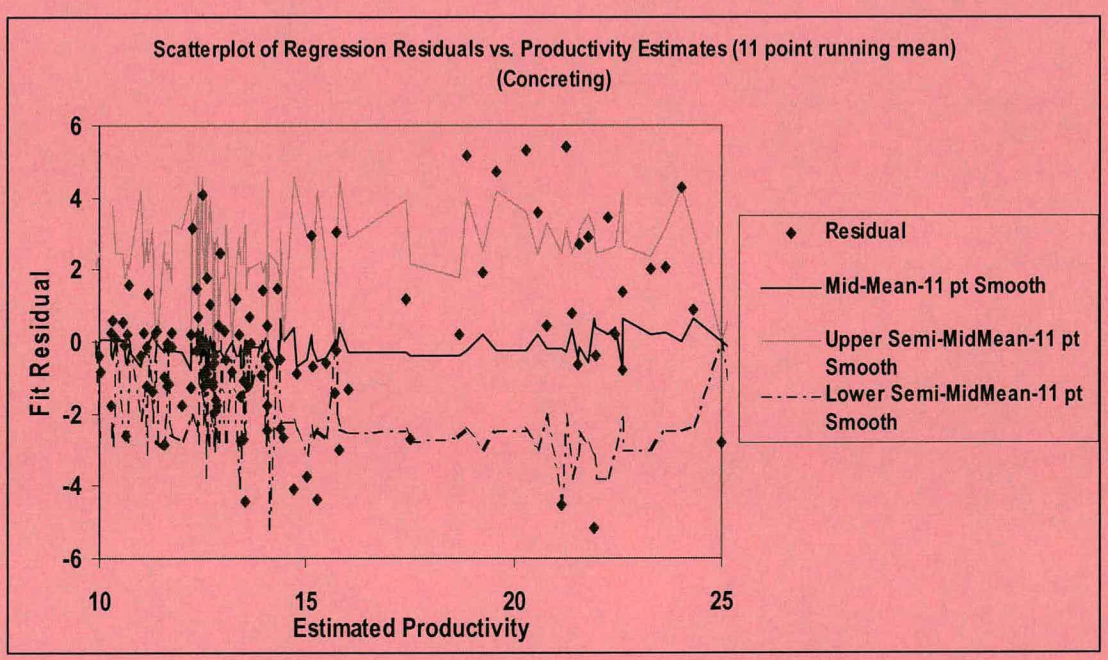


Figure C.2: Linear model residuals vs. productivity estimates with statistical measures

- **Examination of scatterplots of residuals against the explanatory variables.** This method is detailed in 5.3.4. In figure C.3, the trend indicates that a linear relationship exists between productivity and project number, as the mean does not vary at all with a change in project number. However, when considering the month of pour variable (figure C.4) a nonlinear relationship is apparent. The captions highlight a statistically significant nonlinear trend - which is bump shaped (between A and B). Further evidence of a slight nonlinear behaviour in the data is provided by figure C.5, which displays a bowl shaped trend.

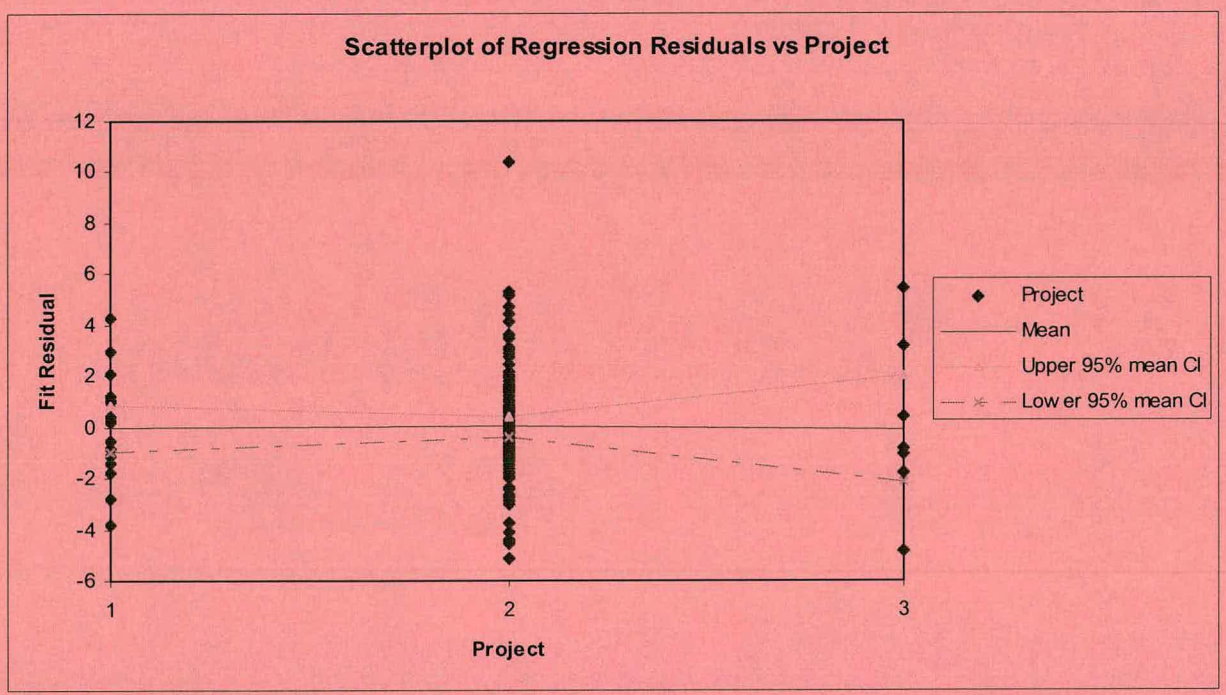


Figure C.3: Scatterplot of productivity linear model residuals vs. the project variable

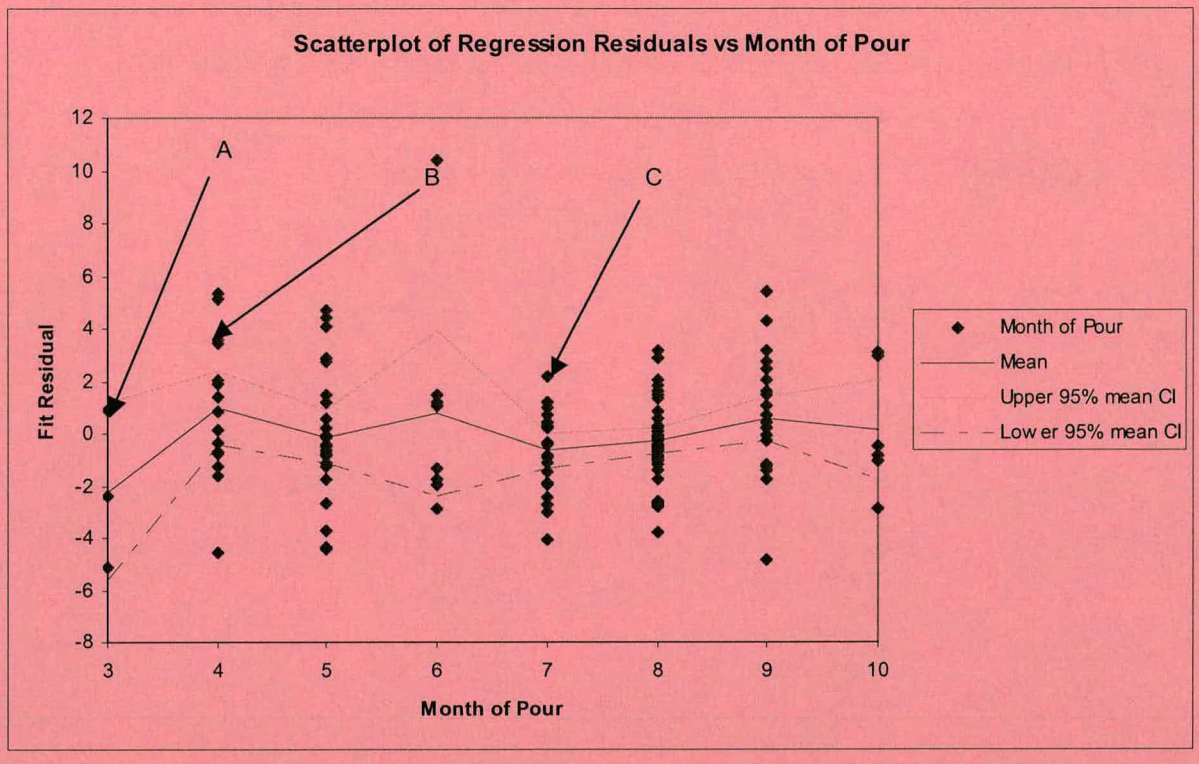


Figure C.4: Productivity linear model residuals vs. the month of pour variable

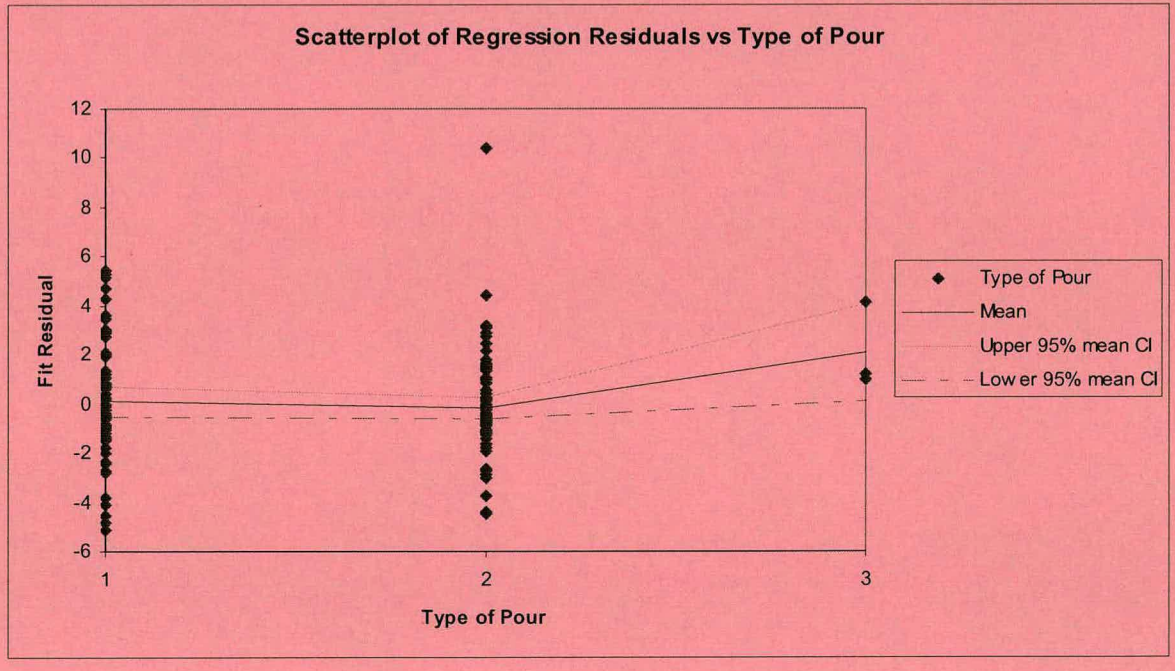


Figure C.5: Productivity linear model residuals vs. the type of pour variable

To conclude, this investigation has discovered that a slight, yet significant nonlinear trend exists in the data, and to answer the above question: the relationship would be described better by a nonlinear model than a linear one. Therefore, only nonlinear modelling methods shall be used to estimate the productivity of the RMC supply process.

APPENDIX D

SOURCE CODE OF COMPUTER MODELS

D.1 Introduction

This appendix contains the source code for three of the models developed in this research project: in section two, *MatSim* – a simulation model developed in the Matlab programming language; in section three, *ExcelSim* – a simulation model developed in Microsoft Excel; and in section four, *CBRSim* – a hybrid model of case-based reasoning and simulation developed in Microsoft Excel.

D.2 Source code of *MatSim*

```
%This is a discrete-event simulation model, MatSim, of the concrete
% supply and delivery process.

%get user inputs
format short
Num_Trucks = input('Enter the number of trucks:    ');
Pour_Vol = input('Enter the pour volume:    ');
lognormal = 0;
gamma = 1;
beta = 2;
Distribution = input('Enter the distribution you wish to use: lognormal, gamma, or beta:    ');
Time1 = [];
Num_Cycles1 = [];
Average_Cycle_Time1 = [];
Average_Batch_Time1 = [];
Average_TraveltoSite_Time1 = [];
Average_Position_Time1 = [];
Average_Pump_Time1 =[];
Average_Washout_Time1 = [];
Average_TraveltoBatcher_Time1 = [];
Average_Batch_Queue_Time1 = [];
Average_Batch_Idle_Time1 = [];
Average_Pump_Queue_Time1 = [];
Average_Pump_Idle_Time1 = [];
Average_Washout_Queue_Time1 = [];
Average_Washout_Idle_Time1 = [];
Total_Batch_Queue_Time1 = [];
Total_Batch_Idle_Time1 = [];
Total_Pump_Queue_Time1 = [];
Total_Pump_Idle_Time1 = [];
Total_Washout_Queue_Time1 = [];
Total_Washout_Idle_Time1 = [];
Volume_Placed1 = [];
Truck_Number1 = [];
Truck_Utilisation1 = [];
Batch_Utilisation1 = [];
Pump_Utilisation1 = [];
Washout_Utilisation1 = [];
```

```
Productivity1 = [];
Labour_Cost1 = [];
Plant_Cost1 = [];
Materials_Cost1 = [];
Total_Cost1 = [];

for Truck_Number = 3:Num_Trucks
%declared variables
A = 1;
Cum_Vol = 0;
Batch_Vol = 8;
%Q1- Batch Queue, Q2-Pump Queue, Q3- Washout Queue
Num_In_Q1 = 0;
Num_In_Q2 = 0;
Num_In_Q3 = 0;
%TA_1- Time of arrival@1, TD_1- Time of departure from 1, TA_2- Time of
%arrival @2, TD_2- Time of departure from 2, TA_3- Time of arrival@3, TD_3-
%Time of departure from 3.
TA_1 = 0;
TD_1 = 0;
TA_2 = 0;
TD_2 = 0;
TA_3 = 0;
TD_3 = 0;
TNA_1 = 0;
a = 1;
b = 1;
c = 1;
d = 1;
e = 1;
f = 1;
g = 1;
x = 1;
y = 1;
z = 1;
Num_Cycles = 0 - Num_Trucks;
BatchQT = 0;
PumpQT = 0;
WashoutQT = 0;
BUSY = 1;
IDLE = 0;
B_Idle_Start_Time = 0;
Batch_Idle_Time = 0;
Total_Batch_Idle_Time = 0;
P_Idle_Start_Time = 0;
Pump_Idle_Time = 0;
Total_Pump_Idle_Time = 0;
W_Idle_Start_Time = 0;
Washout_Idle_Time = 0;
```

```
Total_Washout_Idle_Time = 0;
Batch_Status = BUSY;
Pump_Status = IDLE;
Washout_Status = IDLE;
%set 1st array value.
TA_1(a) = 0;
BatchQT(f) = 0;
PumpQT(g) = 0;
WashoutQT(x) = 0;
Total_Batch_QT = 0;
Total_Pump_QT = 0;
Total_Washout_QT = 0;
Total_Batch_Time = 0;
Total_Travelto_Time = 0;
Total_Position_Time = 0;
Total_Pump_Time = 0;
Total_Washout_Time = 0;
Total_Travelfrom_Time = 0;
%handle to lognormal.
lognormal1 = @lognormal;
gamma1 = @gamma;
beta1 = @beta;
%values of all durations of interest.
if Distribution == lognormal
   Batch = feval(lognormal1, 47.5, 23.25, 10);
   Travelto = feval(lognormal1, 17.31, 5.37, 10);
   Position = feval(lognormal1, 5.74, 6.52, 1.27);
   Pump = feval(lognormal1, 8.89,3.28,5);
   Washout = feval(lognormal1, 3, 1, 2);
   Travelfrom = feval(lognormal1, 17.31, 5.37, 5);
elseif Distribution == gamma
   Batch = feval(gamma1, 10, 0.5, 2);
   Travelto = feval(gamma1, 51.93, 0.63, 10);
   Position = feval(gamma1, 3, 1, 2);
   Pump = feval(gamma1, 4,1,2);
   Washout = feval(gamma1, 3, 1, 2);
   Travelfrom = feval(gamma1, 51.93, 0.63, 10);
else
   Batch = feval(beta1, 0.47, 0.58,19,59, 2);
   Travelto = feval(beta1, 1.6,0.89,9.65,21, 3);
   Position = feval(beta1, 0.2, 0.22,2,13, 2);
   Pump = feval(beta1, 4,1,4,6,2);
   Washout = feval(beta1, 3, 1,1,2, 2);
   Travelfrom = feval(beta1, 1.6, 0.89,9.65,21, 3);
end
%set 1st array value.
TD_1(b) = TA_1(a) + BatchQT(f) + Batch;
TA_2(c) = TD_1(b) + Travelto;
TD_2(d) = TA_2(c) + PumpQT(g) + Position + Pump;
```

```
TA_3(y) = TD_2(d) + 2;
TD_3(z) = TA_3(y) + WashoutQT(x) + Washout;
TNA_1(e) = TD_3(z) + Travelfrom;
%control while loop
while Cum_Vol < Pour_Vol
  %discover next event type and time of occurrence.
  if TA_1(a) < TD_1(b)
    k = TA_1(a);
  else k = TD_1(b);

  end
  if TA_2(c) < TD_2(d)
    l = TA_2(c);
  else l = TD_2(d);
  end
  if TA_3(y) < TD_3(z)
    m = TA_3(y);
  else m = TD_3(z);
  end
  if k < l
    n = k;
  else n =l;
  end
  if n < m
    o = n;
  else o = m;
  end
%perform next event.
if o == TA_1(a)
  Time = TA_1(a);
  if Batch_Status == IDLE
    Batch_Status = BUSY;
    Batch_Idle_Time = Time - B_Idle_Start_Time;
    Total_Batch_Idle_Time = Total_Batch_Idle_Time + Batch_Idle_Time;
  end
  a = a + 1;
  if a < Num_Trucks
  TA_1(a) = TA_1(a-1)+2;
  else TA_1(a) = TNA_1(e);
  end
  if TA_1(a) < TNA_1(e)
    TA_1(a) = TA_1(a);
  else TA_1(a) = TNA_1(e);
  end
    h = TD_1(b) - TA_1(a);
    %if h>0, truck must join batch queue.
    if h > 0
      f = f + 1;
```

```
        %adjust batch queue length.
        if Num_In_Q1 >= Num_Trucks
          Num_In_Q1 = 1;
        else Num_In_Q1 = Num_In_Q1 + 1;
        end
        BatchQT(f) = h;
        Total_Batch_QT = Total_Batch_QT + BatchQT(f);
      end

elseif o == TD_1(b)
  Time = TD_1(b);
  b = b + 1;
  if Distribution == lognormal
    Batch = feval(lognormal1, 47.5, 23.25, 10);
  elseif Distribution == gamma
    Batch = feval(gamma1, 10, 0.5, 2);
  else
    Batch = feval(beta1, 1, 2,19,59, 2);
  end

  Total_Batch_Time = Total_Batch_Time + Batch;
  TD_1(b) = TA_1(b) + BatchQT(f) + Batch;
  %adjust batch queue length.
  if Num_In_Q1 > 0
    Num_In_Q1 = Num_In_Q1 - 1;
  else
    Num_In_Q1 = 0;
    Batch_Status = IDLE;
    B_Idle_Start_Time = Time;
  end
elseif o == TA_2(c)
  Time = TA_2(c);
  if Pump_Status == IDLE
    Pump_Status = BUSY;
    Pump_Idle_Time = Time - P_Idle_Start_Time;
    Total_Pump_Idle_Time = Total_Pump_Idle_Time + Pump_Idle_Time;
  end
  c = c + 1;
  if Distribution == lognormal
    Travelto = feval(lognormal1, 17.35, 5.37, 10);
  elseif Distribution == gamma
    Travelto = feval(gamma1, 51.93, 0.63, 10);
  else
    Travelto = feval(beta1, 1,2,19,59, 3);
  end

  Total_Travelto_Time = Total_Travelto_Time + Travelto;
  TA_2(c) = TD_1(c) + Travelto;
  p = TD_2(d) - TA_2(c);
```

```
    %if p>0, truck must join pump queue.
    if p > 0
      g = g + 1;

      %adjust pump queue length.
      if Num_In_Q2 > Num_Trucks
        Num_In_Q2 = 1;
      else Num_In_Q2 = Num_In_Q2 + 1;
      end
      PumpQT(g) = p;
      Total_Pump_QT = Total_Pump_QT + PumpQT(g);
    end

elseif o == TD_2(d)
    Time = TD_2(d);
    Cum_Vol = Cum_Vol + Batch_Vol;
    %TA_1(a) = TD_1(b);
    d = d + 1;
    if Distribution == lognormal
      Position = feval(lognormal1, 5.74, 6.52, 10);
      Pump = feval(lognormal1, 8.89,3.28,5);
    elseif Distribution == gamma
      Position = feval(gamma1, 3, 1, 2);
      Pump = feval(gamma1, 4,1,2);
    else
      Position = feval(beta1, 1, 2,2,13, 2);
      Pump = feval(beta1, 4,1,4,6,2);
    end
    Total_Position_Time = Total_Position_Time + Position;
    Total_Pump_Time = Total_Pump_Time + Pump;
    TD_2(d) = TA_2(d) + PumpQT(g) + Position + Pump;
    %adjust pump queue length.
    if Num_In_Q2 > 0
      Num_In_Q2 = Num_In_Q2 - 1;
    else
      Num_In_Q2 = 0;
      Pump_Status = IDLE;
      P_Idle_Start_Time = Time;
    end
elseif o == TA_3(y)
    Time = TA_3(y);
    if Washout_Status == IDLE
      Washout_Status = BUSY;
      Washout_Idle_Time = Time - W_Idle_Start_Time;
      Total_Washout_Idle_Time = Total_Washout_Idle_Time + Washout_Idle_Time;
    end
    y = y + 1;
    %TA_1(a) = TD_1(b);
    TA_3(y) = TD_2(d) + 2;
```

```
  q = TD_3(z) - TA_3(y);
  %if q>0, truck must join washout queue.
  if q > 0
    x = x + 1;
    %adjust washout queue length.
    if Num_In_Q3 > Num_Trucks
      Num_In_Q3 = 1;
    else Num_In_Q3 = Num_In_Q3 + 1;
    end
    WashoutQT(x) = q;
    Total_Washout_QT = Total_Washout_QT + WashoutQT(x);
  end
elseif o == TD_3(z)
  Time = TD_3(z);
  Num_Cycles = Num_Cycles + 1;
  %TA_1(a) = TD_1(b);
  z = z + 1;
  if Distribution == lognormal
    Washout = feval(lognormal1, 3, 1, 1);
  elseif Distribution == gamma
    Washout = feval(gamma1, 3, 1, 2);
  else
    Washout = feval(beta1, 3, 1,1,2, 2);
  end

  Total_Washout_Time = Total_Washout_Time + Washout;
  TD_3(z) = TA_3(z) + WashoutQT(x) + Washout;
  e = e + 1;
  if Distribution == lognormal
    Travelfrom = feval(lognormal1, 17.31, 5.37, 10);
  elseif Distribution == gamma
    Travelfrom = feval(gamma1, 51.93, 0.63, 10);
  else
    Travelfrom = feval(beta1, 1.6, 0.89,9.65,21, 3);
  end
  Total_Travelfrom_Time = Total_Travelfrom_Time + Travelfrom;
  TNA_1(e) = TD_3(e) + Travelfrom;
  %adjust washout queue length.
  if Num_In_Q3 > 0
    Num_In_Q3 = Num_In_Q3 - 1;
  else
    Num_In_Q3 = 0;
    Washout_Status = IDLE;
    W_Idle_Start_Time = Time;
  end
 end
end

Time1 = [Time1 Time];
```

```
Num_Cycles1 = [Num_Cycles1 Num_Cycles];
%report results.
Average_Cycle_Time = (Time(1,A)/Num_Cycles(1,A));
Average_Cycle_Time1 = [Average_Cycle_Time1 Average_Cycle_Time];
Average_Batch_Time = (Total_Batch_Time(1,A)/Num_Cycles(1,A));
Average_Batch_Time1 = [Average_Batch_Time1 Average_Batch_Time];
Average_TraveltoSite_Time = (Total_Travelto_Time(1,A)/Num_Cycles(1,A));
Average_TraveltoSite_Time1 = [Average_TraveltoSite_Time1 Average_TraveltoSite_Time];
Average_Position_Time = (Total_Position_Time(1,A)/Num_Cycles(1,A));
Average_Position_Time1 = [Average_Position_Time1 Average_Position_Time];
Average_Pump_Time = (Total_Pump_Time(1,A)/Num_Cycles(1,A));
Average_Pump_Time1 =[Average_Pump_Time1 Average_Pump_Time];
Average_Washout_Time = (Total_Washout_Time(1,A)/Num_Cycles(1,A));
Average_Washout_Time1 = [Average_Washout_Time1 Average_Washout_Time];
Average_TraveltoBatcher_Time = (Total_Travelfrom_Time(1,A)/Num_Cycles(1,A));
Average_TraveltoBatcher_Time1 = [Average_TraveltoBatcher_Time1
Average_TraveltoBatcher_Time];
Average_Batch_Queue_Time = Total_Batch_QT(1,A)/Num_Cycles(1,A);
Average_Batch_Queue_Time1 = [Average_Batch_Queue_Time1
Average_Batch_Queue_Time];
Average_Batch_Idle_Time = Total_Batch_Idle_Time(1,A)/Num_Cycles(1,A);
Average_Batch_Idle_Time1 = [Average_Batch_Idle_Time1 Average_Batch_Idle_Time];
Average_Pump_Queue_Time = Total_Pump_QT(1,A)/Num_Cycles(1,A);
Average_Pump_Queue_Time1 = [Average_Pump_Queue_Time1
Average_Pump_Queue_Time];
Average_Pump_Idle_Time = Total_Pump_Idle_Time(1,A)/Num_Cycles(1,A);
Average_Pump_Idle_Time1 = [Average_Pump_Idle_Time1 Average_Pump_Idle_Time];
Average_Washout_Queue_Time = Total_Washout_QT(1,A)/Num_Cycles(1,A);
Average_Washout_Queue_Time1 = [Average_Washout_Queue_Time1
Average_Washout_Queue_Time];
Average_Washout_Idle_Time = Total_Washout_Idle_Time(1,A)/Num_Cycles(1,A);
Average_Washout_Idle_Time1 = [Average_Washout_Idle_Time1
Average_Washout_Idle_Time];
Total_Batch_Queue_Time = Total_Batch_QT(1,A);
Total_Batch_Queue_Time1 = [Total_Batch_Queue_Time1 Total_Batch_Queue_Time];
Total_Batch_Idle_Time = Total_Batch_Idle_Time(1,A);
Total_Batch_Idle_Time1 = [Total_Batch_Idle_Time1 Total_Batch_Idle_Time];
Total_Pump_Queue_Time = Total_Pump_QT(1,A);
Total_Pump_Queue_Time1 = [Total_Pump_Queue_Time1 Total_Pump_Queue_Time];
Total_Pump_Idle_Time = Total_Pump_Idle_Time(1,A);
Total_Pump_Idle_Time1 = [Total_Pump_Idle_Time1 Total_Pump_Idle_Time];
Total_Washout_Queue_Time = Total_Washout_QT(1,A);
Total_Washout_Queue_Time1 = [Total_Washout_Queue_Time1
Total_Washout_Queue_Time];
Total_Washout_Idle_Time = Total_Washout_Idle_Time(1,A);
Total_Washout_Idle_Time1 = [Total_Washout_Idle_Time1 Total_Washout_Idle_Time];
%Time;
%Complete_Pumping_Time = Time
Volume_Placed = Cum_Vol(1,A);
```

```
Volume_Placed1 = [Volume_Placed1 Volume_Placed];
%Number_Of_Cycles = Num_Cycles
Truck_Number1 = [Truck_Number1 Truck_Number(1,A)];
%Number_Of_Trucks_Used = Num_Trucks
Truck_Utilisation = ((Time(1,A) - (Total_Batch_QT(1,A) + Total_Pump_QT(1,A) +
Total_Washout_QT(1,A)))/Time(1,A))*100;
Truck_Utilisation1 = [Truck_Utilisation1 Truck_Utilisation];
Batch_Utilisation = ((Time(1,A) - Total_Batch_Idle_Time(1,A))/Time(1,A))*100;
Batch_Utilisation1 = [Batch_Utilisation1 Batch_Utilisation];
Pump_Utilisation = ((Time(1,A) - Total_Pump_Idle_Time(1,A))/Time(1,A))*100;
Pump_Utilisation1 = [Pump_Utilisation1 Pump_Utilisation];
Washout_Utilisation = ((Time(1,A) - Total_Washout_Idle_Time(1,A))/Time(1,A))*100;
Washout_Utilisation1 = [Washout_Utilisation1 Washout_Utilisation];
Productivity = (Volume_Placed(1,A)*60/Time(1,A));
Productivity1 = [Productivity1 Productivity];
Labour_Cost = (Time(1,A)/60)*(9.66*1+7.15*5+6.82*2+9.16*1);
Labour_Cost1 = [Labour_Cost1 Labour_Cost];
Plant_Cost = (Time(1,A)/60)*(20.08*1+18.00*Num_Trucks(1,A)+0.89*3);
Plant_Cost1 = [Plant_Cost1 Plant_Cost];
Materials_Cost = Volume_Placed(1,A)*57.84;
Materials_Cost1 = [Materials_Cost1 Materials_Cost];
Total_Cost = Labour_Cost(1,A) + Plant_Cost(1,A) + Materials_Cost(1,A);
Total_Cost1 = [Total_Cost1 Total_Cost];

A = A+1;
end

Truck_Number1
Total_Cost1
Productivity1
%bar graph of cost versus truck number
subplot(2,2,1); bar(Truck_Number1, Total_Cost1);
xlabel('Number of Trucks')
ylabel('Total Cost')
Title('Cost v Number of Trucks v Completion Time')
overlap_plot(Truck_Number1, Time1);
ylabel('Operation Completion Time')
subplot(2,2,2); bar(Truck_Number1, Productivity1)
xlabel('Number of Trucks')
ylabel('Productivity')
Title('Productivity v Number of Trucks')

subplot(2,2,3); plot(Truck_Number1, Truck_Utilisation1, Truck_Number1, Batch_Utilisation1,
Truck_Number1,Pump_Utilisation1, Truck_Number1,Washout_Utilisation1)
xlabel('Number of Trucks')
ylabel('Utilisation')
legend('Truck', 'Batch', 'Pump', 'Washout', -1)
Title('Utilisations v Number of Trucks')
n = 2;
```

```
figure(n);
costgraph = bar(Truck_Number1, Total_Cost1);
xlabel('Number of Trucks')
ylabel('Total Cost')
Title('Cost v Number of Trucks')
overlap_plot(Truck_Number1, Time1);
ylabel('Operation Completion Time')

n = n + 1;
figure(n);
prodgraph = bar(Truck_Number1, Productivity1);
xlabel('Number of Trucks')
ylabel('Productivity')
Title('Productivity v Number of Trucks')

n = n + 1;
figure(n);
plot(Truck_Number1, Truck_Utilisation1, Truck_Number1, Batch_Utilisation1,
Truck_Number1,Pump_Utilisation1, Truck_Number1,Washout_Utilisation1)
xlabel('Number of Trucks')
ylabel('Utilisation')
legend('Truck', 'Batch', 'Pump', 'Washout', -1)
Title('Utilisations v Number of Trucks')
```

D.3 Source code of *ExcelSim*

% The model *ExcelSim* is very different to the model *MatSim*. *ExcelSim* is based %upon 1 VBA macro, which automatically performs required actions on a series of %spreadsheets when the macro is invoked. Whereas, *MatSim* is based upon pure %programming code.

% Due to the nature of software development in Excel it is extremely difficult to provide an adequate description of an entire program using words. Thus, a copy of the model, *ExcelSim* has been provided with this thesis.

%Below is the macro, Replications, which runs the program.

```
Sub Replications()
StartTrucks = Application.InputBox("Enter Start Number of Trucks")
EndTrucks = Application.InputBox("Enter End Number of Trucks")
NumberOfReplications = Application.InputBox("Enter number of Replications (1000 Max)", ,
1000, , , , 1)

For Experiments = StartTrucks To EndTrucks
   ' Clear Output Sheet
   Sheets("Output").Select
   Rows("2:1002").Select
   Range("A1002").Activate
   Selection.ClearContents
   Range("A2").Select
```

```
  Sheets("Simulation Sheet").Select
  Cells(2, 2).Value = Experiments

  ' ExperimentID = Application.InputBox("Enter Experiment ID", , "Exp 2(iv)")
  Application.ScreenUpdating = False
  Sheets("Summary").Select
  Cells(6, 6).Value = NumberOfReplications
  ' Sheets("Input Data").Select
  ' Cells(5, 6).Value = ExperimentID
  For Output = 2 To NumberOfReplications + 1
    Randomize
    Sheets("Simulation Sheet").Select
    Rows("45:45").Select
    Selection.Copy
    Sheets("Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
      :=False, Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Simulation Sheet").Select
    Application.CutCopyMode = False
    Range("j71").Select
      Selection.AutoFill Destination:=Range("j71:j6"), Type:=xlFillDefault
    Range("m71").Select
      Selection.AutoFill Destination:=Range("m71:m6"), Type:=xlFillDefault
    Range("r71").Select
      Selection.AutoFill Destination:=Range("r71:r6"), Type:=xlFillDefault
    Range("u71").Select
      Selection.AutoFill Destination:=Range("u71:u6"), Type:=xlFillDefault
    Range("Ab71").Select
      Selection.AutoFill Destination:=Range("Ab71:Ab6"), Type:=xlFillDefault
    Range("Ad71").Select
      Selection.AutoFill Destination:=Range("Ad71:Ad6"), Type:=xlFillDefault
  Next Output
  Sheets("Simulation Sheet").Select
  Cells(1, 1).Select
  Sheets("Summary").Select
  Application.ScreenUpdating = True
  ActiveWorkbook.SaveAs Filename:=Experiments, FileFormat:=xlNormal, Password:="",
WriteResPassword:="", _
    ReadOnlyRecommended:=False, CreateBackup:=False
Next Experiments
End Sub
```

D.4 Source code of *CBRSim*

% The model *CBRSim* is based %upon 2 VBA macros: one to invoke the CBR element of the model; and, one to invoke simulation. These macros automatically perform the required actions on a series of %spreadsheets when invoked.

% To reiterate from above, due to the nature of software development in Excel it is extremely difficult to provide an adequate description of an entire program using words. Thus, a copy of the model, *CBRSim* has been provided with this thesis.

%Below is the CBR macro, CBR_Threshold, which selects the probability distributions that represent the *times of interest*, for use in simulation.

```
Sub CBR_Threshold()
%retrieves cases from CBR_Sim_CaseBase sheet if the matching factor is over 75%.
%The cases are presented in a summary table on CBR_Output sheet, for selection by the user.
Dim i As Integer
Dim j As Integer
Dim k As Integer
Dim n As Integer
Dim o As Integer
Dim p As Integer
Dim q As Integer
Dim r, s, t, u As Integer
   Sheets("CBR_Output").Select
   Rows("2:348").Select
   Selection.ClearContents
   Range("D350").Select
   Sheets("User_Input").Select
   Range("BL51:BY128").Select
   Selection.ClearContents
   Range("BL51").Select
Dim myString As String
j = 2
k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 46).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 46).Select
ActiveCell.Offset(0, -39).Select
   ActiveCell.Range("A1:AD1").Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 6).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 45).Select
   Selection.Copy
   Sheets("User_Input").Select
   Cells(k, 65).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
```

```
    Cells(i, 1).Select
    Selection.Copy
    Sheets("CBR_Output").Select
    Cells(j, 5).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
    Sheets("User_Input").Select
    Cells(k, 64).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
j = j + 1
k = k + 1
End If
Next i
k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 53).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 53).Select
ActiveCell.Offset(0, -46).Select
    ActiveCell.Range("A1:AD1").Select
    Selection.Copy
    Sheets("CBR_Output").Select
    Cells(j, 6).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
    Sheets("CBR_Sim_CaseBase").Select
    Cells(i, 52).Select
    Selection.Copy
    Sheets("User_Input").Select
    Cells(k, 67).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
    Sheets("CBR_Sim_CaseBase").Select
    Cells(i, 1).Select
    Selection.Copy
    Sheets("CBR_Output").Select
    Cells(j, 5).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
    Sheets("User_Input").Select
    Cells(k, 66).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
j = j + 1
k = k + 1
End If
```

```
Next i
k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 55).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 55).Select
ActiveCell.Offset(0, -48).Select
   ActiveCell.Range("A1:AD1").Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 6).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 54).Select
   Selection.Copy
   Sheets("User_Input").Select
   Cells(k, 69).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 1).Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 5).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("User_Input").Select
   Cells(k, 68).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
j = j + 1
k = k + 1
End If
Next i

k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 57).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 57).Select
ActiveCell.Offset(0, -50).Select
   ActiveCell.Range("A1:AD1").Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 6).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
```

```
      False, Transpose:=False
    Sheets("CBR_Sim_CaseBase").Select
    Cells(i, 56).Select
    Selection.Copy
    Sheets("User_Input").Select
    Cells(k, 71).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
    Sheets("CBR_Sim_CaseBase").Select
    Cells(i, 1).Select
    Selection.Copy
    Sheets("CBR_Output").Select
    Cells(j, 5).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
    Sheets("User_Input").Select
    Cells(k, 70).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
j = j + 1
k = k + 1
End If

Next i


k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 59).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 59).Select
ActiveCell.Offset(0, -52).Select
   ActiveCell.Range("A1:AD1").Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 6).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 58).Select
   Selection.Copy
   Sheets("User_Input").Select
   Cells(k, 73).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 1).Select
   Selection.Copy
```

```
    Sheets("CBR_Output").Select
    Cells(j, 5).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
        False, Transpose:=False
    Sheets("User_Input").Select
    Cells(k, 72).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
        False, Transpose:=False
j = j + 1
k = k + 1
End If

Next i

k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 61).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 61).Select
ActiveCell.Offset(0, -54).Select
    ActiveCell.Range("A1:AD1").Select
    Selection.Copy
    Sheets("CBR_Output").Select
    Cells(j, 6).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
        False, Transpose:=False
    Sheets("CBR_Sim_CaseBase").Select
    Cells(i, 60).Select
    Selection.Copy
    Sheets("User_Input").Select
    Cells(k, 75).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
        False, Transpose:=False
    Sheets("CBR_Sim_CaseBase").Select
    Cells(i, 1).Select
    Selection.Copy
    Sheets("CBR_Output").Select
    Cells(j, 5).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
        False, Transpose:=False
    Sheets("User_Input").Select
    Cells(k, 74).Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
        False, Transpose:=False
j = j + 1
k = k + 1
End If
```

```
Next i

k = 51
For i = 2 To 68
n = Sheets("CBR_Sim_CaseBase").Cells(i, 63).Value
If n = "1" Then
Sheets("CBR_Sim_CaseBase").Select
Cells(i, 63).Select
ActiveCell.Offset(0, -56).Select
   ActiveCell.Range("A1:AD1").Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 6).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 62).Select
   Selection.Copy
   Sheets("User_Input").Select
   Cells(k, 77).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("CBR_Sim_CaseBase").Select
   Cells(i, 1).Select
   Selection.Copy
   Sheets("CBR_Output").Select
   Cells(j, 5).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
   Sheets("User_Input").Select
   Cells(k, 76).Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, SkipBlanks:= _
      False, Transpose:=False
j = j + 1
k = k + 1
End If
Next i

Sheets("User_Input").Select
Range("bs78").Select

o = Sheets("CBR_Sim_CaseBase").Range("AT86").Value
p = Sheets("CBR_Sim_CaseBase").Range("BA86").Value
q = Sheets("CBR_Sim_CaseBase").Range("BC86").Value
r = Sheets("CBR_Sim_CaseBase").Range("BE86").Value
s = Sheets("CBR_Sim_CaseBase").Range("BG86").Value
t = Sheets("CBR_Sim_CaseBase").Range("BG86").Value
u = Sheets("CBR_Sim_CaseBase").Range("BG86").Value
If o = 0 And p = 0 And q = 0 And r = 0 And s = 0 And t = 0 And u = 0 Then
```

```
myString = MsgBox("There isn't an available matching case, enter the distribution manually
before running simulation.", vbOKOnly)
Range("AV12").Select
End If

End Sub
```

%Below is the simulation macro, Run_Simulation, which uses the probability distributions that were selected by CBR to recreate the RMC supply process.

```
Sub Run_Simulation()
Dim Output, a, b, c, x, y, z, e, f, g, h
a = 3
b = Sheets("User_Input").Range("M2").Value
NoOfReps = Sheets("User_Input").Range("AF2").Value

  Sheets("1Truck_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
  Selection.ClearContents
  Range("A2").Select
  Sheets("2Trucks_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
  Selection.ClearContents
  Range("A2").Select
  Sheets("3Trucks_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
  Selection.ClearContents
  Range("A2").Select
  Sheets("4Trucks_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
  Selection.ClearContents
  Range("A2").Select
  Sheets("5Trucks_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
  Selection.ClearContents
  Range("A2").Select
  Sheets("6Trucks_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
  Selection.ClearContents
  Range("A2").Select
  Sheets("7Trucks_Output").Select
  Rows("2:1002").Select
  Range("A1002").Activate
```

```
    Selection.ClearContents
    Range("A2").Select
    Sheets("8Trucks_Output").Select
    Rows("2:1002").Select
    Range("A1002").Activate
    Selection.ClearContents
    Range("A2").Select
    Sheets("9Trucks_Output").Select
    Rows("2:1002").Select
    Range("A1002").Activate
    Selection.ClearContents
    Range("A2").Select
    Sheets("10Trucks_Output").Select
    Rows("2:1002").Select
    Range("A1002").Activate
    Selection.ClearContents
    Range("A2").Select
    Sheets("11Trucks_Output").Select
    Rows("2:1002").Select
    Range("A1002").Activate
    Selection.ClearContents
    Range("A2").Select
    Sheets("12Trucks_Output").Select
    Rows("2:1002").Select
    Range("A1002").Activate
    Selection.ClearContents
    Range("A2").Select

k = 6
Do While e <= f
    e = Sheets("Truck_Simulation").Cells(k, 26).Value
    f = Sheets("User_Input").Range("AD2").Value
    k = k + 1
Loop
l = 57
Do While g <= h
    g = Sheets("Truck_Simulation").Cells(l, 26).Value
    h = Sheets("User_Input").Range("AD2").Value
    l = l + 1
Loop
m = 108
Do While i <= j
    i = Sheets("Truck_Simulation").Cells(m, 26).Value
    j = Sheets("User_Input").Range("AD2").Value
    m = m + 1
Loop
n = 159
Do While e1 <= f1
    e1 = Sheets("Truck_Simulation").Cells(n, 26).Value
```

```
    f1 = Sheets("User_Input").Range("AD2").Value
    n = n + 1
Loop
o = 210
Do While g1 <= h1
    g1 = Sheets("Truck_Simulation").Cells(o, 26).Value
    h1 = Sheets("User_Input").Range("AD2").Value
    o = o + 1
Loop
p = 261
Do While i1 <= j1
    i1 = Sheets("Truck_Simulation").Cells(p, 26).Value
    j1 = Sheets("User_Input").Range("AD2").Value
    p = p + 1
Loop
q = 312
Do While e2 <= f2
    e2 = Sheets("Truck_Simulation").Cells(q, 26).Value
    f2 = Sheets("User_Input").Range("AD2").Value
    q = q + 1
Loop
r = 363
Do While g2 <= h2
    g2 = Sheets("Truck_Simulation").Cells(r, 26).Value
    h2 = Sheets("User_Input").Range("AD2").Value
    r = r + 1
Loop
s = 414
Do While i2 <= j2
    i2 = Sheets("Truck_Simulation").Cells(s, 26).Value
    j2 = Sheets("User_Input").Range("AD2").Value
    s = s + 1
Loop
t = 465
Do While e3 <= f3
    e3 = Sheets("Truck_Simulation").Cells(t, 26).Value
    f3 = Sheets("User_Input").Range("AD2").Value
    t = t + 1
Loop
u = 516
Do While g3 <= h3
    g3 = Sheets("Truck_Simulation").Cells(u, 26).Value
    h3 = Sheets("User_Input").Range("AD2").Value
    u = u + 1
Loop
v = 567
Do While e4 <= f4
    e4 = Sheets("Truck_Simulation").Cells(v, 26).Value
    f4 = Sheets("User_Input").Range("AD2").Value
```

```
    v = v + 1
Loop

If b = 1 Then
For Output = 2 To NoOfReps + 1
    Sheets("Truck_Simulation").Select
    Rows(k - 2).Select
    Selection.Copy
    Sheets("1Truck_Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks:=False,
Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False

    Randomize
    Range("J617").Select
      Selection.AutoFill Destination:=Range("J617:J6"), Type:=xlFillDefault
    Range("M617").Select
      Selection.AutoFill Destination:=Range("M617:M6"), Type:=xlFillDefault
    Range("R617").Select
      Selection.AutoFill Destination:=Range("R617:R6"), Type:=xlFillDefault
    Range("U617").Select
      Selection.AutoFill Destination:=Range("U617:U6"), Type:=xlFillDefault
    Range("AB617").Select
      Selection.AutoFill Destination:=Range("AB617:AB6"), Type:=xlFillDefault
    Range("AD617").Select
      Selection.AutoFill Destination:=Range("AD617:AD6"), Type:=xlFillDefault

Next Output

ElseIf b = 2 Then
For Output = 2 To NoOfReps + 1
    Sheets("Truck_Simulation").Select
    Rows(l - 2).Select
    Selection.Copy
    Sheets("2Trucks_Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks:=False,
Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False

    Randomize
    Range("J617").Select
      Selection.AutoFill Destination:=Range("J617:J6"), Type:=xlFillDefault
    Range("M617").Select
      Selection.AutoFill Destination:=Range("M617:M6"), Type:=xlFillDefault
    Range("R617").Select
```

```
    Selection.AutoFill Destination:=Range("R617:R6"), Type:=xlFillDefault
  Range("U617").Select
    Selection.AutoFill Destination:=Range("U617:U6"), Type:=xlFillDefault
  Range("AB617").Select
    Selection.AutoFill Destination:=Range("AB617:AB6"), Type:=xlFillDefault
  Range("AD617").Select
    Selection.AutoFill Destination:=Range("AD617:AD6"), Type:=xlFillDefault

Next Output
End If

For Output = 2 To NoOfReps + 1
  For c = a To b
    If c = "3" Then
      Sheets("Truck_Simulation").Select
      Rows(m - 2).Select
      Selection.Copy
      Sheets("3Trucks_Output").Select
       Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
      :=False, Transpose:=False
      Cells(Output, 1).Value = Output - 1
      Sheets("Truck_Simulation").Select
      Application.CutCopyMode = False
    ElseIf c = "4" Then
      Sheets("Truck_Simulation").Select
      Rows(n - 2).Select
      Selection.Copy
      Sheets("4Trucks_Output").Select
      Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
      :=False, Transpose:=False
      Cells(Output, 1).Value = Output - 1
      Sheets("Truck_Simulation").Select
      Application.CutCopyMode = False
    ElseIf c = "5" Then
      Sheets("Truck_Simulation").Select
      Rows(o - 2).Select
      Selection.Copy
      Sheets("5Trucks_Output").Select
      Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
      :=False, Transpose:=False
      Cells(Output, 1).Value = Output - 1
      Sheets("Truck_Simulation").Select
      Application.CutCopyMode = False
    ElseIf c = "6" Then
      Sheets("Truck_Simulation").Select
      Rows(p - 2).Select
      Selection.Copy
      Sheets("6Trucks_Output").Select
      Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
```

```
    :=False, Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False
  ElseIf c = "7" Then
    Sheets("Truck_Simulation").Select
    Rows(q - 2).Select
    Selection.Copy
    Sheets("7Trucks_Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
    :=False, Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False
  ElseIf c = "8" Then
    Sheets("Truck_Simulation").Select
    Rows(r - 2).Select
    Selection.Copy
    Sheets("8Trucks_Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
    :=False, Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False
  ElseIf c = "9" Then
    Sheets("Truck_Simulation").Select
    Rows(s - 2).Select
    Selection.Copy
    Sheets("9Trucks_Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
    :=False, Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False
  ElseIf c = "10" Then
    Sheets("Truck_Simulation").Select
    Rows(t - 2).Select
    Selection.Copy
    Sheets("10Trucks_Output").Select
    Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
    :=False, Transpose:=False
    Cells(Output, 1).Value = Output - 1
    Sheets("Truck_Simulation").Select
    Application.CutCopyMode = False
  ElseIf c = "11" Then
    Sheets("Truck_Simulation").Select
    Rows(u - 2).Select
    Selection.Copy
    Sheets("11Trucks_Output").Select
```

```
        Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False
        Cells(Output, 1).Value = Output - 1
        Sheets("Truck_Simulation").Select
        Application.CutCopyMode = False
      ElseIf c = "12" Then
        Sheets("Truck_Simulation").Select
        Rows(t - 2).Select
        Selection.Copy
        Sheets("12Trucks_Output").Select
        Rows(Output).PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False
        Cells(Output, 1).Value = Output - 1
        Sheets("Truck_Simulation").Select
        Application.CutCopyMode = False
      End If
    Next c
      Randomize
    Range("J617").Select
      Selection.AutoFill Destination:=Range("J617:J6"), Type:=xlFillDefault
    Range("M617").Select
      Selection.AutoFill Destination:=Range("M617:M6"), Type:=xlFillDefault
    Range("R617").Select
      Selection.AutoFill Destination:=Range("R617:R6"), Type:=xlFillDefault
    Range("U617").Select
      Selection.AutoFill Destination:=Range("U617:U6"), Type:=xlFillDefault
    Range("AB617").Select
      Selection.AutoFill Destination:=Range("AB617:AB6"), Type:=xlFillDefault
    Range("AD617").Select
      Selection.AutoFill Destination:=Range("AD617:AD6"), Type:=xlFillDefault
  Next Output
  Sheets("Output_Summary").Select
  Range("A1").Select


  End Sub
```

APPENDIX E

CALCULATION OF THE PROCESS PERFORMANCE MEASURES

E.1 Introduction

This appendix contains the details of how the process performance measures are calculated by the simulation models, *MatSim* and *ExcelSim* and the hybrid model, *CBRSim*. Upon the calculation of these measures a report is produced by the report generator subprogram and is presented to the user.

The process performance measures (and how they are calculated) are:

Process duration (hours) = value of the simulation clock when the stopping condition is met, measured in hours (E.1)

Direct cost (pounds) = unit cost of truck mixers (pounds/hour) x number of truck mixers x process duration (hours) + unit cost of employees (pounds/hour) x number of employees x process duration (hours) + unit cost of RMC (pounds/m^3) x volume discharged (m^3) (E.2)

$$\textbf{Productivity (m3/hour)} = \frac{total\ volume\ discharged\ (m^3)}{process\ duration\ (hours)} \tag{E.3}$$

Truck mixer utilisation rate (%) =

$$100 \times (1 - \frac{queue1\ idle\ time + queue2\ idle\ time + queue3\ idle\ time}{process\ duration}) \tag{E.4}$$

$$\textbf{Server utilisation rate (\%)} = 100 \times (1 - \frac{server\ idle\ time}{process\ duration}) \tag{E.5}$$

APPENDIX F

FIT OF PROBABILITY DISTRIBUTIONS TO THE TIMES OF INTEREST

F.1 Introduction

This appendix contains the outcome of the assessment of how well the 27 probability distributions (detailed in chapter 6) fitted the times of interest data. This assessment was made visually and using formal goodness-of-fit tests, the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test. The results presented in graphs F.1 to F.5 show the percentage of times that a particular distribution was found to provide the best fit to the data, by the 3 assessment methods mentioned above.

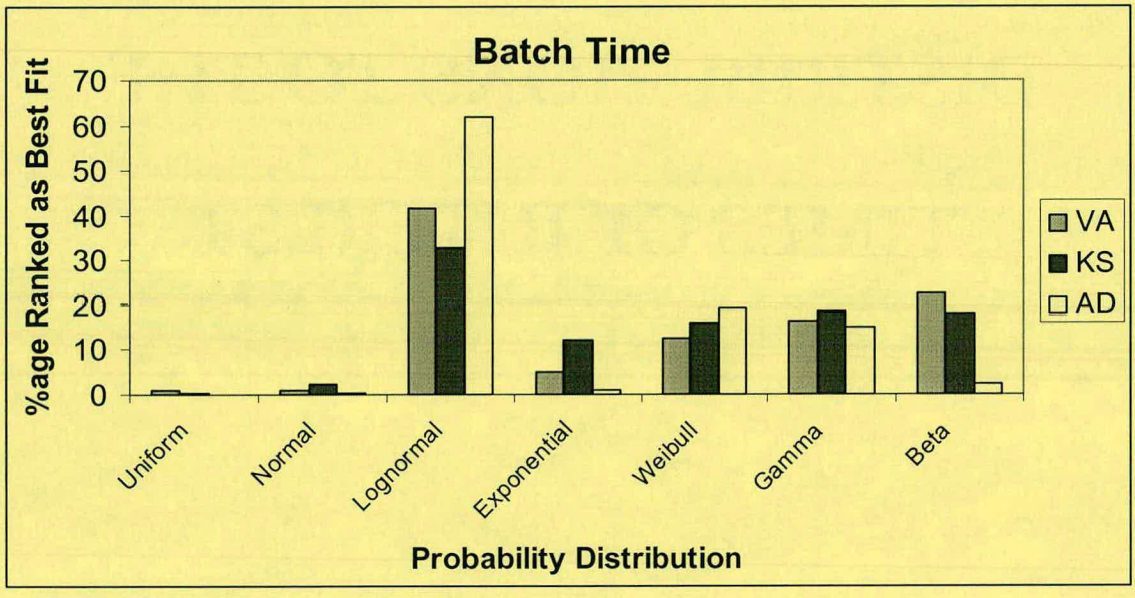


Figure F.1: Assessment of fit of PDFs to the batch time data

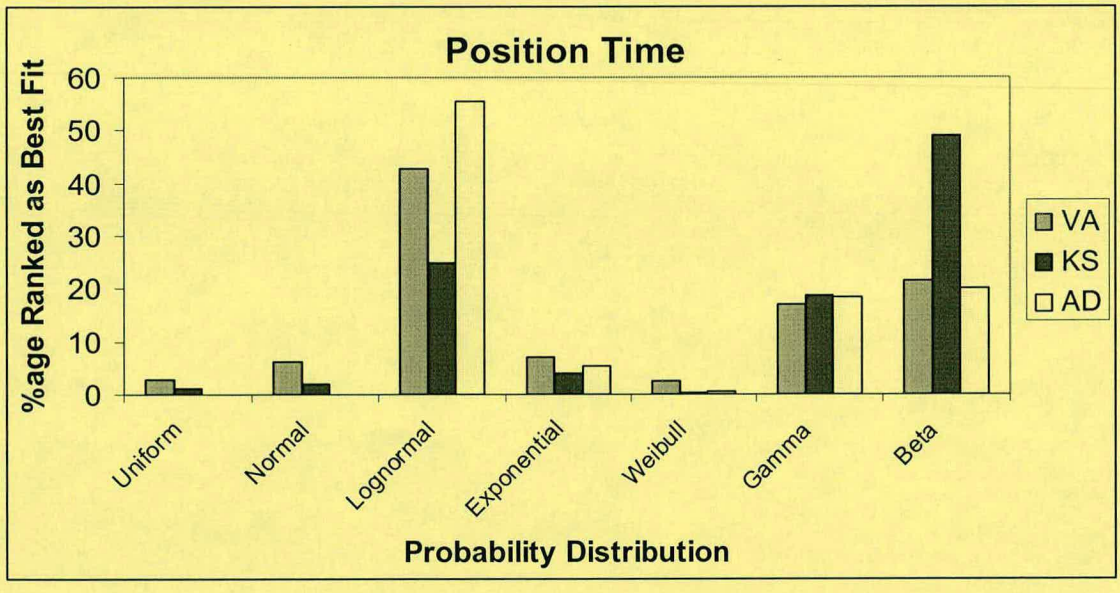


Figure F.2: Assessment of fit of PDFs to the position time data

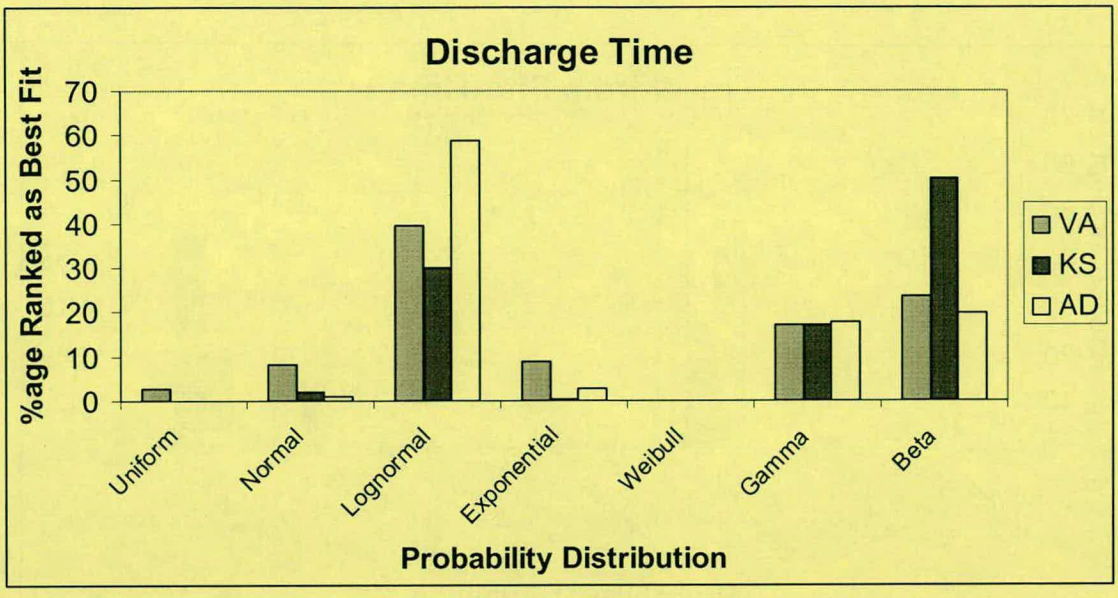


Figure F.3: Assessment of fit of PDFs to the discharge time data

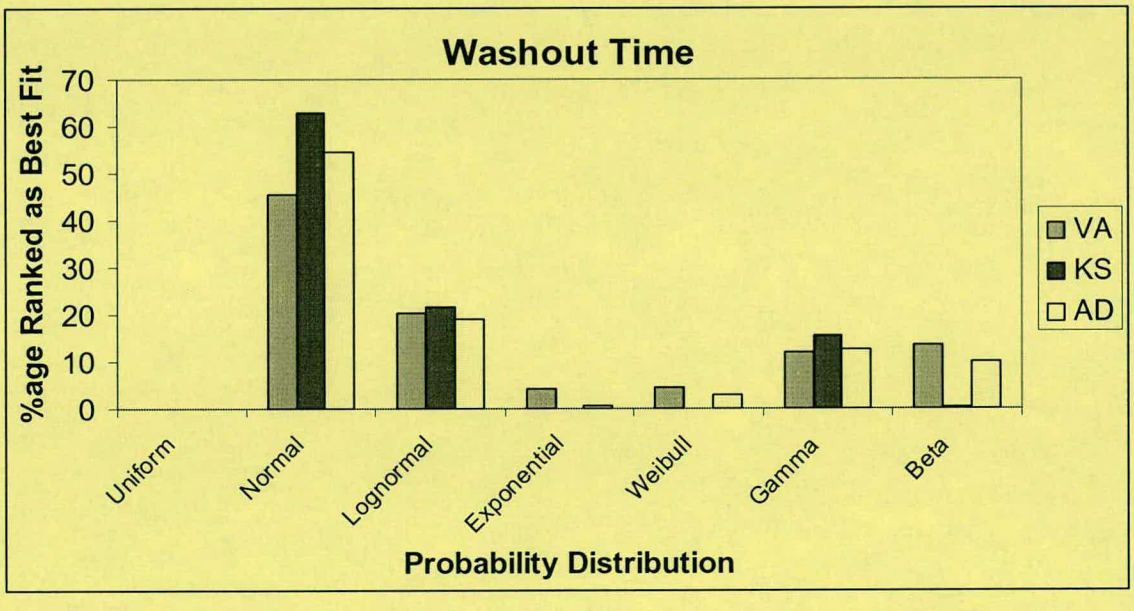


Figure F.4: Assessment of fit of PDFs to the washout time data

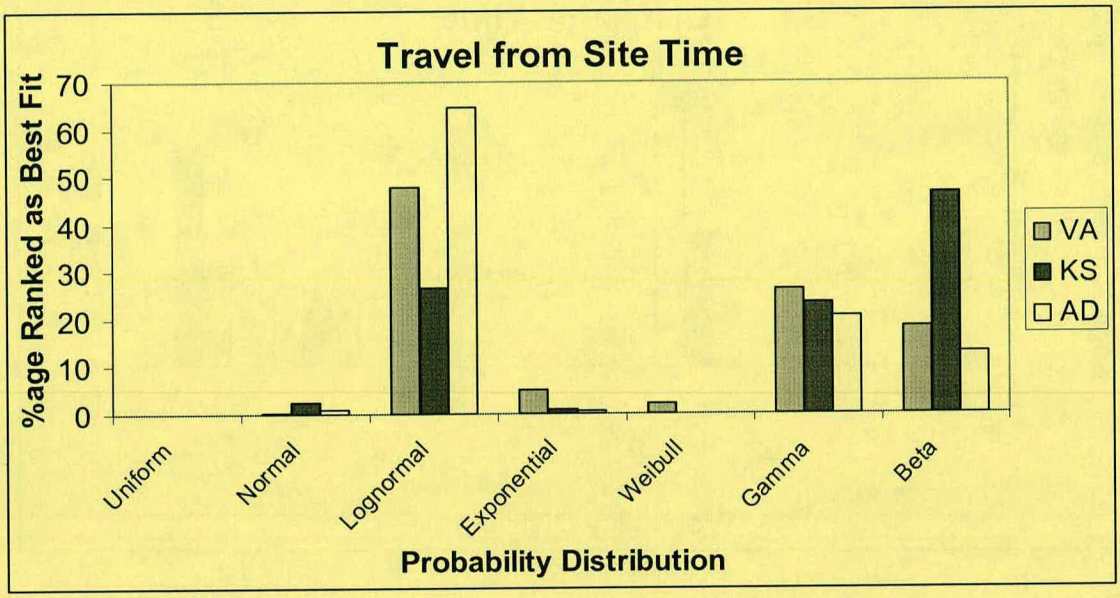


Figure F.5: Assessment of fit of PDFs to the travel from site time data

APPENDIX G

VALIDATION OF *ExcelSim*

G.1 Introduction

The *Microsoft Excel*-based model, *ExcelSim* was validated using the same method (correlated inspection approach) and data (observed operations) as those used in the case of *MatSim* as detailed in 6.5. Additionally, the probability distributions (and their parameters) that were chosen by the author to be used in the *MatSim* model were also used here. This should result in a fair validation process for *ExcelSim* and allow a direct comparison to be made between the two models.

A summary of the correlated inspection approach to the validation of *ExcelSim* is shown in figure G.1. It is notable that the errors are of a similar magnitude to those produced by *MatSim* (from figure 6.7). Thus the *ExcelSim* model can be declared as valid as that model.

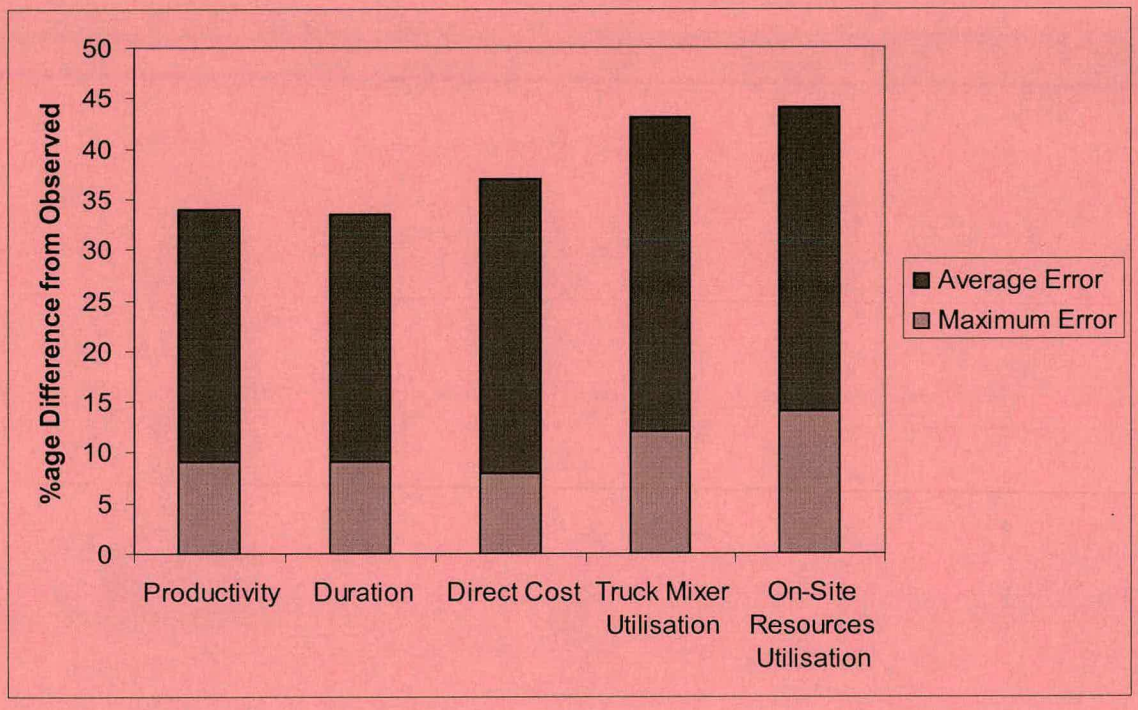


Figure G.1: Correlated inspection results from the validation of *ExcelSim*

APPENDIX H

MATHEMATICAL NOTATION

H.1 Introduction

This appendix contains the notation of a select list of equations that were presented in this thesis. They are as follows:

Equation 5.1: Mahalanobis distance measure $D^2 = (x-\mu)' V^{-1} (x-\mu)$

where: D^2 is the Mahalanobis distance measure; x = vector of data; μ = vector of mean values of explanatory variables; and V^{-1} = inverse covariance matrix of explanatory variables.

Equation 5.2: Penrose distance measure $P_{ij} = \sum_{k=1}^{p} \frac{(\mu_{ki} - \mu_{kj})^2}{pV_k}$

where: P is the Penrose distance measure; i and j are separate populations; p is the number of degrees of freedom of the data or the number of explanatory variables; μ = vector of mean values of explanatory variables; and V = variance of explanatory variables.

Equation 6.4: Maximum likelihood estimator $L(\theta) = f_\theta(X_1) \bullet f_\theta(X_2) \cdots f_\theta(X_n)$

where: $L(\theta)$ is the likelihood function, which is just the joint probability mass function since the data are independent; and $f_\theta(X_n)$ denotes the probability mass function for the particular probability distribution being fitted to the data.

Equation 6.7: Kolmogorov-Smirnov test $D_n = \sup_x \{ |F_n(x) - F(x)| \}$

where: $F_n(x)$ is the empirical distribution function; F(x) is the fitted distribution function; and the sup of a set of numbers A is the smallest value that is greater than or equal to all members of A.

Equation 6.8: Anderson-Darling test $A_n^2 = n \int_{-\infty}^{\infty} \left[F_n(x) - F(x) \right]^2 \Psi(x) f(x) dx$

where: $\Psi(x)$ is a weighting function that ensures more influence is attributed to the tails (right and left) of the underlying data histogram than the centre; and $f(x)$ denotes the probability mass function for the particular probability distribution being fitted to the data.

APPENDIX I

PUBLICATIONS

I.1 Introduction

This appendix contains the full versions of the papers produced during the course of this research project (as of 30th June 2005). The following journal papers have been included (with publication details[4]):

1. Graham, D., Smith, S. D. and Crapper, M. (2004). "Improving concrete placement simulation with a case-based reasoning input." *Civil Engineering and Environmental Systems*, 21(2), 137-150.
2. Graham, D. and Smith, S. D. (2004). "Estimating the productivity of cyclic construction operations using case-based reasoning." *Advanced Engineering Informatics*, 18, 17-28.
3. Graham, D., Smith, S. D. and Dunlop, P. (2005). "Lognormal distribution provides an optimal representation of the concrete delivery and placement process." *Journal of Construction Engineering and Management*, ASCE, 131(2), 230-238.
4. Graham, D., Forbes, D. R. and Smith, S. D. (2004)."Modeling the ready mixed concrete delivery system with neural networks." *Automation in Construction*, accepted for publication.
5. Graham, D., Smith, S. D., and Tommelein, I. D. (2005)."In-situ concreting: identification and quantification of contractor costs." *Construction Management and Economics*, under review.

The following peer-reviewed conference papers have been included (with publication details):

1. Graham, D., and Smith, S.D. (2003). "A hybrid model to improve the estimation of concreting operations." *ARCOM 19th annual conference*. University of Brighton, Sept 3-5, 505-512.
2. Graham, D., and Smith S.D. (2004). "A method for effectively implementing construction process productivity estimation models." *ARCOM 20th annual conference*, Heriot-Watt University, Sept 1-3, 1043-1052.
3. Forbes, D.R, McGurnaghan, H., Graham, D., and Smith S.D. (2004). "Concrete placing productivity using a novel neural network design." *ARCOM 20th annual conference*, Heriot-Watt University, Sept 1-3, 1053-1062.
4. Graham, D., and Smith, S.D. (2005). "A model to facilitate the optimization of the concrete supply process." *Construction Research Congress, ASCE*, San Diego, April 5-7, 58-62.

[4] Permission to re-print these publications has been granted by their respective publishers.

5. Graham, D., Smith, S. D. and Tommelein, I. D. (2005). "Cost of concrete placement for contractors." *ARCOM 21st annual conference*, University of Salford, Sept 5-7, accepted for publication.
6. Graham, D., Tommelein, I. D. and Smith, S. D. (2005). "Cost reduction through vertical integration of the in-situ concrete supply chain." *ARCOM 21st annual conference*, University of Salford, Sept 5-7, accepted for publication.

Civil Engineering and Environmental Systems
Vol. 21, No. 2, June 2004, pp. 137–150

Spon Press
Taylor & Francis Group

IMPROVING CONCRETE PLACEMENT SIMULATION WITH A CASE-BASED REASONING INPUT

DARREN GRAHAM, SIMON D. SMITH* and MARTIN CRAPPER

School of Engineering and Electronics, University of Edinburgh, Alexander Graham Bell Building, The Kings Buildings, Edinburgh, EH9 3JL, United Kingdom

(Received 18 March 2003; In final form 24 February 2004)

Discrete-event simulation has been used extensively in the past to analyse construction operations and has been shown to be an effective tool for improving construction process planning. Unfortunately, the widespread application of simulation has been prevented, in part, by the requirement for the developer and user to understand the stochastic features of the process. Further, if the stochastic inputs are not representative of the real system, the simulation output will be misleading. This article proposes that case-based reasoning (CBR) could be used to: improve the input and output data of a simulation model and remove the requirement for a user to have specialist statistical knowledge of the process. Case-based reasoning works by solving new cases from knowledge stored in a case base.

Two models are presented in this article: a discrete-event simulation model, MatSim, and a hybrid CBR-simulation model, CBRSim. Both models were developed using real construction data. The models were compared to measure any improvements in simulation output, by using a CBR suggested input. Data from an independent construction project were used to test both models, and the results indicate that CBRSim can achieve estimates of observed values that are more accurate and reliable than those from MatSim.

Keywords: Concrete placement; Stochastic processes; Discrete-event simulation Case-based reasoning

INTRODUCTION

Simulation has been shown as an effective tool for improving construction process planning [1], but unfortunately, usage difficulties have prevented a wide application of the technology in industry [2]. For instance, constructing and using a simulation model requires an understanding of the stochastic features of the process concerned [3]. While these features are usually and usefully represented by probability distributions, the importance of their effect on the accuracy of simulation output cannot be overemphasised [4, 5]. Within these contexts, this article presents an artificial intelligence technique, case-based reasoning (CBR) as a method to: reduce the level of statistical knowledge required to use simulation; and to improve the input to, and consequently the output of, a simulation model.

The concrete supply, delivery and placement process is stochastic in nature and requires analysis using non-deterministic techniques. One such technique is discrete-event simulation, which has been used to analyse construction operations for over three decades [6].

* Corresponding author. E-mail: simon.smith@ed.ac.uk

ISSN 1028-6608 print; ISSN 1029-0249 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/10286600410001688937

One of the first major computer simulation packages for construction operations was CYCLONE, developed by Halpin [7]. Since its introduction, construction simulation has advanced with computer technology [8]. Various construction simulation languages have been developed, and Naji and Najafi suggested that there are two main simulation methodologies: network-based and graphical [9]. Opdenbosch and Baker criticised network-based simulation claiming it failed to describe three main factors that influence most construction operations: the construction site, the geometry of the building, and the dynamics involved in operating construction equipment [10]. Due to these reasons, the graphical simulation technique is gaining support within the construction industry [9]. DISCO [11] and COOPS [12] provide graphical interfaces to allow the user to construct a simulation model by manipulating graphical symbols on a computer screen [8].

Two models of the concrete placing system are presented: a discrete-event simulation model, MatSim and a hybrid CBR-simulation model, CBRSim. Both models are based on probability distributions that have been derived from real construction data. A correlated inspection between the output of the models and observed data, from an independent construction site, is presented as part of the validation process. This indicates that the results obtained using CBR-improved simulation input are more representative of the real system than standard simulation.

STOCHASTIC MODELLING OF THE CONCRETE SUPPLY, DELIVERY AND PLACEMENT PROCESS

In previous research, concreting operations have been separated into two processes: batching and delivery [*e.g.* Refs. 13, 14]; and pumping and placing [*e.g.* Refs. 15, 16]. In this study, the concrete supply, delivery and placing process is considered in its entirety, and is shown in Figure 1. In an ideal situation, a truck mixer is filled with concrete at the batching plant and travels to the construction site. On arrival, the truck mixer moves

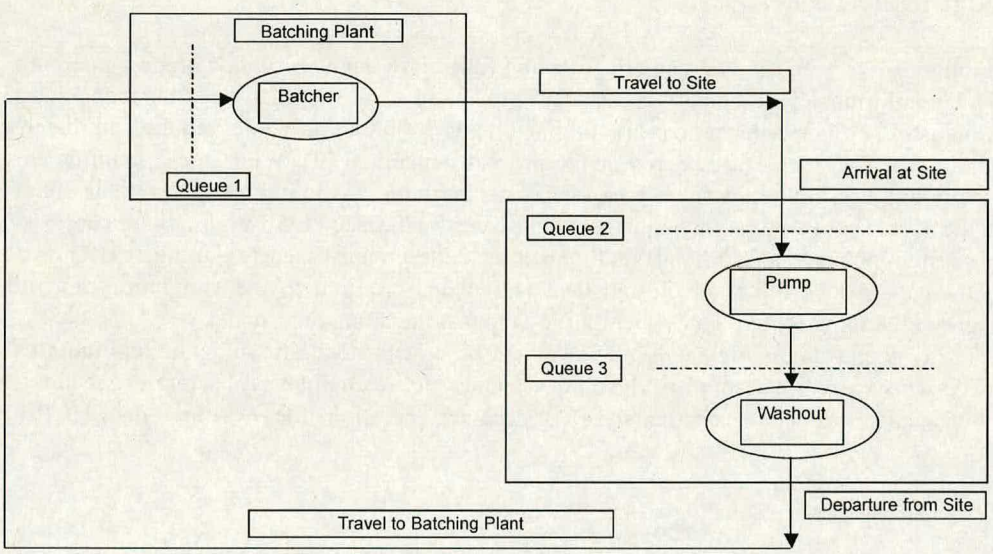


FIGURE 1 The concrete supply, delivery and placement process.

into position at the pump and discharges its load. Finally, the empty truck mixer is cleaned in the washout area and proceeds back to the batching plant, thereby completing the cycle. However, as the concreting process is stochastic, the events that occur within the system (*e.g.* truck arrival times and pump start times) take place at irregular intervals [15]. Queuing of trucks at the batching plant, pump and washout areas can be expected, as it is often unlikely that a truck will arrive simultaneous to the previous truck departing. If trucks arrive late, there will be a lengthening of the process and under-utilisation of the plant and labour. To create a realistic simulation model, it is necessary to reproduce the unpredictability of the process, and this can be done through using probability distributions. The ideal distributions that represent the times of interest of the process (*e.g.* travel time to site) are identified from real construction data. These distributions can produce times, which are used to recreate a typical cycle of the process. Through the simulation of multiple cycles, the attributes of a particular operation set-up can be provided, such as pour duration and truck or pump utilisation rates. To ensure that the model is realistic and accurate, it must be based on data collected from real construction operations.

DATA COLLECTION

The data used to build both the simulation model, MatSim and the hybrid model, CBRSim, were collected during the time study of four construction projects in Scotland, UK. These projects involved the construction of wastewater treatment plants. In total, 220 concrete pours were observed and the variables of interest were recorded. These variables were pour date and type, weather, target and actual slump, batch time, arrival on site time, position time, start and completion of pumping times, washout time, batch quantity, total pour volume, truck wait time (at batcher and pump), batcher and pump idle times and the number of trucks involved with the pour.

Upon completion of this exercise, the data were entered into a spreadsheet and the times of interest were extracted. In the case of MatSim, these times were used to identify a theoretical probability distribution, representative of the concreting process. In CBRSim, the theoretical probability distribution, which ideally represents each time of interest for each pour, were recorded and used in the model.

The data, which were used to validate both models, were collected from a construction project independent from those detailed above. The project took place in the North of England, United Kingdom and involved the construction of a motorway viaduct. From this set of data, 20 pours were selected at random, for use in the validation process.

DISCRETE-EVENT SIMULATION MODELLING

MatSim is a discrete-event simulation model, developed in the Matlab programming environment (the Mathworks, Inc). The following steps were followed in the development of MatSim:

(a) Choice of probability distribution for individual time components.
(b) Generation of random samples ('variates') from these distributions.
(c) Simulation of concreting operations.

Fitting Probability Distributions to Data

The aim of fitting probability distributions to real construction data is to identify a distribution that is an acceptable representation of the concreting process. This study has used a commercially available package, BestFit by Palisade Corporation, to fit theoretical probability distributions to observed data. BestFit calculates the parameters of 28 distributions fitted to these data, using maximum likelihood estimators (MLEs), and provides a plot of the underlying data histogram against probability density function (PDF; Fig. 2). In this study, a visual appraisal was undertaken to determine the optimum distribution, as this is often the superior method of assessing the fit [17]. Generally, it was observed that the beta, gamma and lognormal distributions represented the process appropriately. Therefore, any subsequent investigation into the validity of the MatSim model should consider results produced from variates of all three distributions. Table I shows the parameters, which were estimated for the lognormal, gamma and beta distributions.

Generation of Random Variates

The general method of generating random variates is to transform a uniformly distributed random number (between 0 and 1) through an algorithm, into a series of numbers that correspond to the particular distribution required. In MatSim, the beta and lognormal variates are generated using methods detailed in Law and Kelton [18] and Devroye [19], respectively. The generation of gamma random variates was undertaken using an acceptance–rejection technique, the GB algorithm, developed by Cheng [20].

Running the Simulation Model

The MatSim program is based on three FIFO single-server queues, working in series. These queues are representative of real queues occurring at the batching plant, pump and washout area. In this system, the truck mixers are customers and the batcher, pump and washout are servers. When an event occurs, such as the arrival of a truck on site, the system is updated. As the system evolves over time, trucks depart and others arrive at the pump queue, and the frequency with which these events occur determines the number of loads of concrete delivered [15]. As MatSim proceeds, other statistics such as pump idle time or the truck

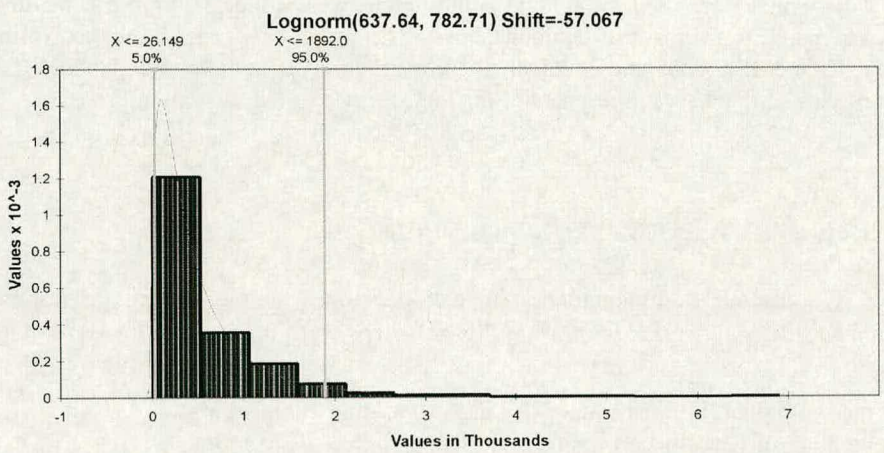


FIGURE 2 An example of PDF *vs.* input data histogram plot from BestFit.

TABLE I Estimated parameters for the lognormal, gamma and beta distributions.

	Distribution							
	Lognormal		*Gamma*		*Beta*			
Time	μ	σ	α	β	α_1	α_2	a	b
Batch	30	6	2	5	1.2	0.58	0	30
Travel	6.74	2.92	1.48	8	1.6	0.89	9.6	21
Position	5.28	6.16	1.48	3	1.4	0.66	2	28
Pump	6.65	1.79	4.28	3.2	1.6	0.75	5	12
Washout	5	2.3	2	3	1	0.66	1	10

wait time in each queue may be recorded. From these times, values of interest such as pump utilisation and truck utilisation can be calculated. MatSim is run from an event list, which contains the next time when each type of event will occur [18]. This time is a sample from the chosen distribution, produced by the random variate generator. For instance, in MatSim, the arrival time of a truck on site is generated by addition of the time of departure from the batching plant plus a sample of the travel time distribution. Finally, simulation will be completed when the user-specified concrete pour volume has been placed.

The MatSim model is used to investigate different concreting operations. Each operation will have its own set of attributes. To allow this, the chosen distribution must be defined over a wide range of mean times. This is done by assuming that the shape of the distribution remains constant, over all ranges of time. For example, when using the lognormal distribution, the shape parameter, σ remains constant (See Tab. I), but the scale parameter, μ must be calculated using the MLE method. On establishing the correct parameters, simulation may now be undertaken.

CASE-BASED REASONING

Case-based reasoning is an artificial intelligence technique, whose origins lie in the work of Shanks and Abelson [21]. It is the process of reasoning and learning by storing cases – records of specific prior reasoning episodes – and retrieving and adapting them to aid new problem solving or interpretation in similar situations [22]. The construction domain is a growing area for applications of CBR, and there have been numerous systems created to address issues in the area of design [23]. These include: CASETOOL, a system for bridge design [24]; and, BRIDGER, a system to aid the conceptual design of cable-stayed bridges [25]. In the area of planning, Tah *et al.* created CBRidge Planner to aid the planning of highway bridge projects [23], and Dzeng and Tommelein developed CasePlan for scheduling and planning boiler erection [26]. All of these CBR systems design, diagnose or plan, but they do not predict. In the past, CBR has been used successfully in solving prediction and estimation problems [27], and an application of CBR to predict construction litigation outcome was researched by Arditi and Tokdemir [28]. This study intends to utilise CBR to predict the optimum distributions that represent a particular construction operation. The predicted distributions will subsequently be used in a simulation study. Aamodt and Plaza described the CBR process as cyclical, comprising the following [29]:

- Retrieval of the most similar case(s), from the case base.
- Reuse of the case(s) to attempt to solve the problem.

- Revision of the proposed solution, if necessary, to solve the problem.
- Retention of the modified solution as a new case.

CBRSim is based on this cyclical process, and is shown in Figure 3. The following stages were addressed in developing CBRSim:

(a) Case base design
(b) Case retrieval
(c) Validation of CBR System
(d) Simulation using CBRSim.

Case Base Design

The case base is the storage area for prior cases. A robust case library, containing a representative and well-distributed set of cases, is the foundation for a good CBR system [30]. Cases must be parameterised, typically by being divided into a series of distinct features by means of which their similarity to other cases can be judged. In CBRSim, the case base stores 220 cases, all cases sharing the same record structure with 5 input and 5 output features. The input features are the number of trucks, pour volume, pour type, month of pour and weather, and each feature is separated into numerical ranges (Tab. II). There are undoubtedly more input

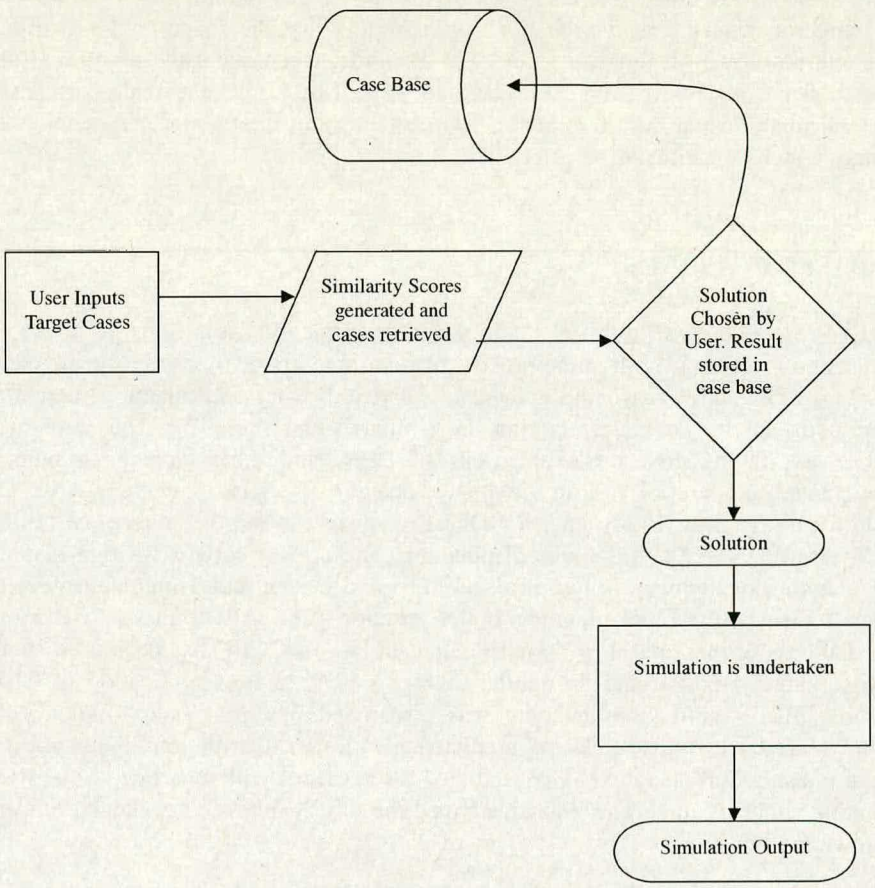


FIGURE 3 Flowchart of CBRSim program.

TABLE II Features and their numeric values, from CBRSim model.

	Feature				
Index number	*Type of pour*	*Month of pour*	*Pour volume*	*Number of trucks*	*Weather*
1	Base	December–February	0–25	1–3	Sunny
2	Wall	March–May	26–50	4–6	Overcast
3	Aeration tank	June–August	51–75	7+	Rain
4	Settlement tank	Septermber–November	76–100		Snow
5	Pipe surround		101–200		Clear
6	Column base		201–250		
7	Floor slab		251+		
8	Roof slab				

features that influence the concreting cycle than the five considered here. However, if an accurate and robust model is produced then the simplicity is justified [31]. The output features are the probability distributions and parameters, which best represent the concreting process. To obtain these features, real construction data was processed using BestFit, in the manner discussed earlier in this article. The next stage of the CBR process is retrieval.

Case Retrieval

A CBR system derives its power from its ability to retrieve relevant cases quickly and accurately from its case base [28], and CBRSim uses two methods to retrieve cases: feature counting, and weighted feature counting. In feature counting, all the features have an equal impact on the outcome. This method has the disadvantage that the analyst cannot give more importance to one feature than another [28]. In weighted feature computation, a weighting is calculated for each feature using the ID3 algorithm, developed by Quinlan [32]. The feature weights calculated for the case base in CBRSim, on its first use, are shown in Table III. As expected the number of trucks and the pour volume are the most influential factors of the system. The weights are updated as the case base grows (*i.e.*) after each run of CBRSim. Both feature counting and weighted feature counting are based on calculating a local similarity measurement, which is subsequently transformed into a measurement of interest using the methods provided by Hand et al. [33]: feature counting, City-Block, Euclidean, Minkowski, weighted City-Block, weighted Minkowski and Maximum-Biased.

On completion of the retrieval process, CBRSim presents the user with a report containing the retrieved case numbers and their corresponding similarity measurements (in percentage form), in descending order of similarity. There is a threshold for similarity, which is 75% in this study, below which a case is excluded from the report. Finally, the user is required to

TABLE III Feature weightings calculated by ID3 algorithm.

Feature	*Weighting*
Type of pour	0.10
Month of pour	0.07
Number of trucks	0.37
Pour volume	0.39
Weather	0.07
Total	1.00

select a case from the list. Frequently, the target case features are not identical to the features of the selected cases. In this situation, CBRSim retains the target case features and the distribution information, as a new case in the case base, thereby completing the cyclic CBR process.

Validation of the CBR System

The retrieval accuracy of the CBR system must be validated before any investigation of the hybrid CBR-simulation model can take place. Validation consists of using the CBR system to determine outcomes of cases where the outcome is already known, to establish that the system produces accurate results. This is typically achieved by removing the output fields from some of the original case base records and determining whether the CBR system is able to put them back again accurately. To carry out validation, Tokdemir developed three methods [34]. Each method uses the similarity measurements generated by the system in a different way to predict the best distributions to use in a simulation study. These are:

- In method 1, the outcome of the retrieved case that received the highest similarity measurement is selected. This method selects cases on the basis of the strength of similarity.
- In method 2, the outcome that appears most frequently is selected from the outcomes of the first 10 cases that have a similarity measurement of 75% or more. In this method, the frequency of cases with high similarity is considered.
- In method 3, the outcome of the cases that received the highest average similarity measurement is selected from among the outcomes of the first five cases that have a similarity measurement of 75% or more.

The case base was split into training and test sets of ratio: 90:10, 80:20 and 70:30, in each case the test set having its output fields removed. The cases to make each set were selected at random. The statistical reliability of the results will be improved by testing the CBR system with differing case base sizes [35]. The test cases provided the input into the CBR system, and the results were analysed using the aforementioned methods and are shown in Table IV.

A value of 100%, using analysis method 1 (See Tab. IV) signifies that the system retrieved at least one case, with a similarity score of 75% or over, for each test case. The CBR system achieved retrieval rates of 100% for all three training/test set ratios. As expected, the results improved as the case base size increased. In general, the prediction rates achieved were better

TABLE IV Results for CBR validation.

Similarity assessment method	*Prediction methods*								
	Training:test ratio (90:10)			*Training:test ratio (80:20)*			*Training:test ratio (70:30)*		
	1	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
Feature counting	88	13	13	82	18	18	68	12	8
City-block	100	50	25	100	59	18	96	44	24
Weighted city-block	100	63	38	100	65	35	96	68	24
Euclidean	100	63	25	100	41	41	100	64	20
Minkowski	100	50	25	100	47	47	96	60	20
Weighted-Minkowski	100	50	25	100	53	53	100	68	20
Maximum-biased	100	63	50	100	59	59	96	40	40

Note: All values in percentage form.

in the 90:10 ratio set, than those for the 70:30 ratio set. The results have provided a sufficient level of validation to justify proceeding to the next stage of the study, investigating the hybrid CBR-simulation model, CBRSim.

Simulation Using CBRSim

The determination of the best case for a given construction operation, ends the CBR side of the hybrid model. The probability distribution parameters, from the selected case, are used by CBRSim to generate random variates for use in simulation. CBRSim can theoretically generate random variates from any of the 28 distributions fitted by BestFit. However, from the results only seven distributions provided representations of the concreting process; uniform, normal, lognormal, exponential, weibull, gamma and beta distributions. The methodology used in CBRSim to undertake simulation of the concreting process is the same as in the MatSim program. However, the computing methods used in the two programs are different. The hybrid model, CBRSim was developed using a spreadsheet package, Microsoft Excel. This allows the user to understand fully each step in the CBR process and simulation of concreting. The next stage of this study is to validate the MatSim and CBRSim programs. The validation process shall also be used to compare the accuracy of the two simulation systems.

VALIDATION OF MODELS

Validation is the process of determining whether a simulation model is an accurate representation of the system, for the particular objectives of a study [18]. Naylor and Finger proposed a three-stage validation procedure that has been adopted by most modellers of both economic and scientific systems [36]. The stages are:

1. Formulation of the hypotheses and postulates for the system under study using the available knowledge such as observation, relevant theory and even intuition. Law and Kelton proposed expanding this stage by considering the face validity of the model [18]. Face validity is whether or not the model appears to be reasonable to those who are knowledgeable about the system.
2. Validation of the assumptions upon which the model is based by subjecting them to empirical testing.
3. To compare the simulation output with real-world observations, 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 upon, are reasonable [36]. This is the most definitive test of a simulation model's validity [18] and therefore is the focus of this study.

The concrete placing model is intended to represent the existing system, not to replace it, but to allow estimations of the system output to be produced. Therefore, the model output and data from the existing system can be compared to ascertain if the model is replicatively valid. Naylor and Finger suggested using classical statistical tests such as analysis of variance and Chi-square test, to determine the goodness-of-fit between the model and the existing system output [36]. The tests are not applicable in this case because simulation output data (for a single time series) is usually highly auto-correlated [18]. Therefore, a different approach must be adopted and in this study, a correlated inspection approach was used to indicate the replicative validity of both models. Correlated inspection at its

most extreme requires that the model be 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 indicated the predictive powers of a model [37]. Therefore, the model will be driven with random variates generated from theoretical probability distributions.

Twenty concrete pours (validation observations) were randomly selected from real-world data, collected from a site independent to that on which the models were developed. The MatSim model was investigated when random variates were generated from the lognormal, gamma and beta distributions. The CBRSim model was studied in a different manner, as the CBR process could theoretically suggest any distribution of the 28 assessed by BestFit. Therefore, the distributions and their parameters are likely to differ for each real concrete pour.

Since random samples from probability distributions are used to drive a simulation model through time, these estimates are just particular realisations of random variables that may have large variances [18]. As a result, the estimates could, in a particular simulation run, differ greatly from the corresponding true characteristics of the model. To reduce this difference, the simulation models, MatSim and CBRSim were replicated 100 times for each concrete pour. The performance parameters of interest: overall pour duration, truck utilisation and pump utilisation, were estimated from the mean of the replicated simulation results. Box plots of the percentage difference between observed and simulated values are shown in Figures 4–6.

MatSim Results

Figure 4 shows that the MatSim model has a capability of estimating the overall pour duration to approximately $\pm 20\%$ of the observed values. The model has a propensity to underestimate this performance parameter. To gain a measure of the reliability of the model, the number of validation observations when the estimation was within $\pm 10\%$ of the observed value, were calculated. In the lognormal simulation run, 7 out of 20 validation observations were in this range. This was improved upon by the gamma and beta distribution simulation runs, where 9 out of 20 validation observations were within the specified range of accuracy.

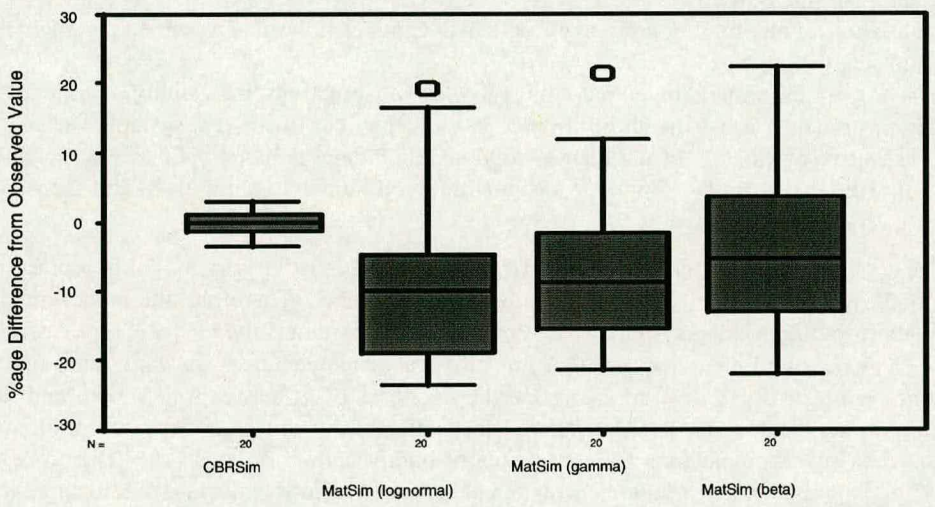


FIGURE 4 Box plots of percentage difference in model and observed values, for overall pour duration.

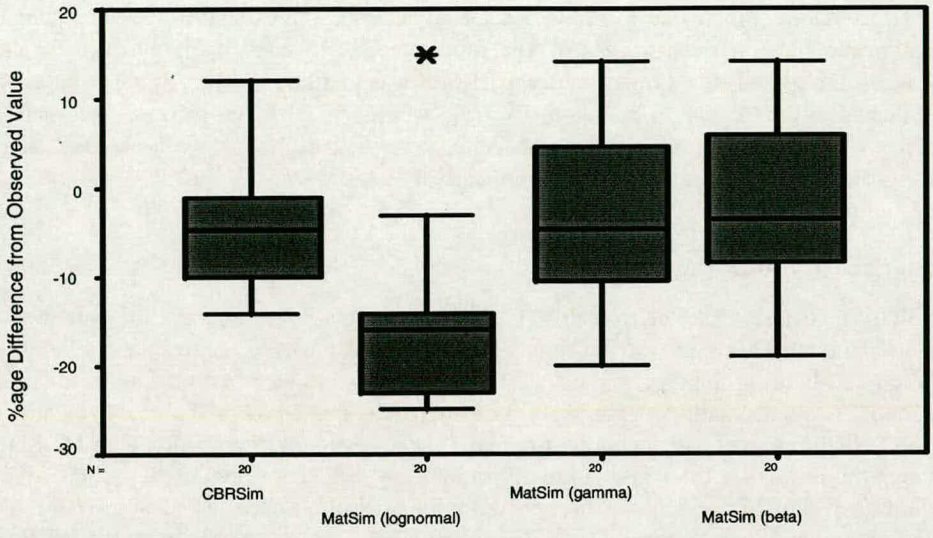


FIGURE 5 Box plots of percentage difference in model and observed values, for truck utilisation.

The truck utilisation rates tend to be underestimated by MatSim (Fig. 5). For this reason, MatSim can be considered a reasonable estimator of this performance parameter – as truck mixers are expensive and it is important to keep utilisation high. In general, MatSim provides a reasonable estimate of the concrete placing system, as the model can provide estimates approximately $\pm 20\%$ of the observed values. The results are sufficiently accurate to tentatively declare MatSim to be a replicatively valid simulation model.

CBRSim Results

The CBRSim model proved itself particularly adept at estimating the pour duration performance parameter (Fig. 4). The difference in the results and observed values ranged from $\pm 3\%$

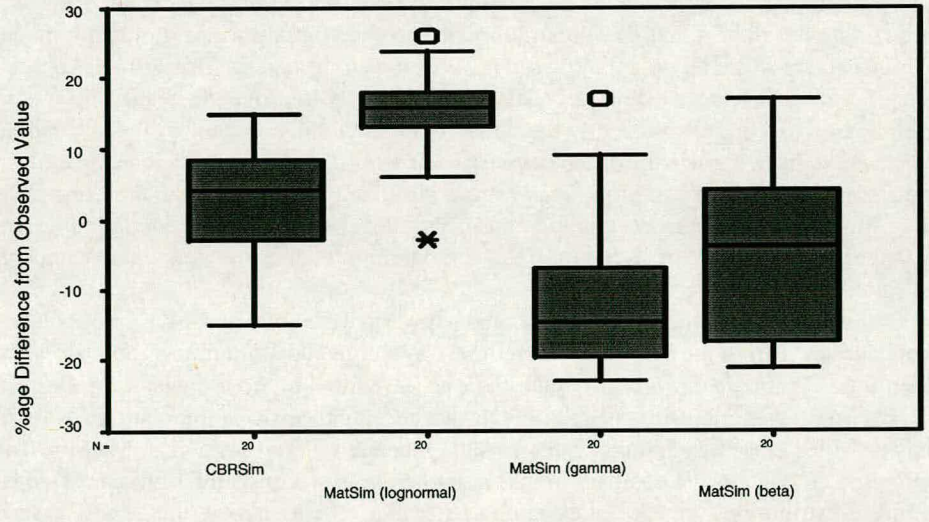


FIGURE 6 Box plots of percentage difference in model and observed values, for pump utilisation.

(Fig. 4). CBRSim was unable to produce such an accurate representation of the resource utilisation parameters (Figs. 5 and 6). The model tended to conservatively estimate the truck utilisation parameter. Consequently, CBRSim was inclined to overestimate the pump utilisation rate (the largest difference is 15%). In summary, CBRSim provides reasonable estimates of the resource utilisation rates and very good estimates of the overall pour duration, and the model can be affirmed replicatively valid.

Comparison of Model Results

The aim of this study is to determine if a CBR derived input could improve the output of a simulation model. This improvement should be measured in terms of accuracy and reliability of the model. Bearing the aim of the study in mind, a direct comparison of the results for each model is appropriate. An examination of the results presented in Figures 4–6, shows that the CBRSim model has the ability to provide better estimates, generally, than MatSim. Although the results for truck and pump utilisations are similar for both models (for certain distributions), CBRSim estimates the pour duration with much greater accuracy (Fig. 4). Further examination of the results, revealed the reliability trends for the models. CBRSim estimated to within 10% of the observed value in 20 out of 20 validation observations for pour duration, 15 out of 20 validation observations for truck utilisation and 16 out of 20 validation observations for pump utilisation. In contrast, MatSim had a poor reliability record for pour duration and pump utilisation, estimating less than 50% of validation observations within 10% of the observed values. The truck utilisation parameter was more consistently calculated when the input was generated from the gamma or beta distributions (Fig. 5).

DISCUSSION AND CONCLUSIONS

The concrete supply, delivery and placement process is random in nature and cannot be analysed using deterministic methods. In the past, operational research techniques such as simulation and neural networks have been used to analyse the concreting process. Simulation has proven to be an effective tool for planning cyclic construction processes, but the widespread use of simulation has been prevented by the requirement to understand the dynamic and stochastic features of the concreting system. Furthermore, the output of the simulation model will be inaccurate if the input is unrepresentative of the environment. This article proposed that an artificial intelligence technique, CBR could primarily improve the quality of input to a simulation model, and consequently the output from the model. Secondly, the user would have no need to have knowledge of the stochastic elements of the system when using CBR to decide upon the input distribution. This article presented two models of the concreting process; MatSim, a discrete-event simulation model, developed using the Matlab programming language; and, CBRSim, a hybrid CBR-simulation model, developed in the Microsoft Excel spreadsheet package.

Data from real construction sites in the United Kingdom were collected. Five times of interest were extracted from the raw data: batch time, travel time, position time, pump time, and washout time. Theoretical probability distributions were fitted to these times, using BestFit. Through visual assessment it was determined that the lognormal, gamma and beta distributions provided close fits, and they were used to generate random variates in MatSim. The CBRSim model contained a database of 220 past operations – consisting of the most suited probability distributions, which represented a particular operational set-up. A new case is compared with each in the database, and the most similar stored case provides the solution

for the new. Therefore, the need for a user to have statistical knowledge of the system has been removed, as the best distribution for a particular operational set-up is chosen by CBR.

A test of the retrieval accuracy of the CBR element of the hybrid model was undertaken. The case base was split into three training/test sets, using the ratios, 90:10, 80:20, and 70:30. Testing on different case base sizes would increase the statistical reliability of the results. The CBR system achieved retrieval rates of 100% for all sets, and in general, these rates improved as the case base size increased.

A validation of both models was undertaken using a correlated inspection approach. The data used for validation was collected from a site independent to those discussed above and 20 pours were selected at random from the site data. The performance parameters investigated were pour duration, truck utilisation and pump utilisation. To improve the statistical significance, simulation was replicated 100 times for each run. The results can be summarised as follows:

- CBRSim achieved estimates of pour duration to within $+3.01/-3.65\%$ of the observed values.
- MatSim achieved an estimate of pour duration of approximately $\pm 20\%$ of the observed values.
- Both models had difficulty in accurately estimating truck and pump utilisation rates. However, in general, truck utilisation was conservatively estimated. This is acceptable as truck mixers are expensive and utilisation must be kept high.
- CBRSim estimated to within $\pm 10\%$ of the observed value in 20 out of 20 validation observations, 15 out of 20 validation observations and 16 out of 20 validation observations, for pour duration, truck utilisation and pump utilisation, respectively. MatSim was unable to match this reliability.

It may be concluded that the MatSim and CBRSim models are useful in planning and estimating concreting operations. From the results, it is clear that using a CBR induced input improves the output of a simulation model. Moreover, CBR removes the need for a user to have statistical knowledge of the concreting process.

References

[1] Halpin D. W., Process-based research to meet the international challenge, *J. Constr. Engrg Mgmt, ASCE*, 1993, 119(3), 415–425.
[2] Shi J., Abourizk S. M., Resource-based modeling for construction simulation, *J. Constr. Engrg Mgmt, ASCE*, 1997, 123(1), 26–33.
[3] Zhang H., Shi J. J., Tam C. M., Iconic animation for activity-based construction simulation, *J. Comp. Civil Engrg, ASCE*, 2002, 16(3), 157–164.
[4] AbouRizk S. M., Halpin D. W., Wilson J. R., Fitting beta distributions based on sample data *J. Constr. Engrg Mgmt, ASCE*, 1994, 120(2), 288–305.
[5] Banks J., Carson J. S., Discrete-Event System Simulation, W. J. Fabrycky, J. H. Mize (editors), Volume 1, Prentice-Hall, Inc, New Jersey, U.S.A., 1984.
[6] Martinez J. C., Ioannou P. G., General-purpose system for effective construction simulation, *J. Constr. Engrg Mgmt*, 1999, 125(4), 265–276.
[7] Halpin D. W., CYCLONE: Method of modeling of job site processes, *J. Constr. Div. ASCE*, 1977, 103, 489–499.
[8] Shi J. J., Activity-based construction (ABC) modelling and simulation method, *J. Constr. Engrg Mgmt, ASCE*, 1999, 125(5), 354–360.
[9] Naji K. K., Najafi F., The future of virtual reality in simulation construction operations *ASCE Annual Conf. Proc.*, 1996.
[10] Opdensbosch A., Baker N., Simulation of Construction Operations in Virtual Interactive Environments, PhD Thesis, Georgia Institute of Technology, U.S.A., 1994.
[11] Huang R., Grigonadis A. M., Halpin D. W., Simulation of cable-stayed bridges using DISCO *Proc. Winter Simulation Conf.*, IEEE, Piscataway, N.J., U.S.A., 1994, 1130–1136.
[12] Liu L. Y., Ioannou P. G., Graphical object-orientated discrete-event simulation system, *Proc, Winter Simulation Conf.*, IEEE, Piscataway, N.J., U.S.A., 1992.

[13] Sawhney A., Abudayyeh O., Chaitavatputtiporn T., Modeling and analysis of concrete production plant using petri nets, *J. Comp. Civil Engrg, ASCE*, 1999, 13(3), 178–186.
[14] Zayed T. M., Halpin D. W., Simulation of concrete batch plant production, *J. Constr. Engrg Mgmt, ASCE*, 2001, 127(2), 132–141.
[15] Smith S. D., Concrete placing analysis using discrete-event simulation, *Proc. Institute of Civ. Engineers, Structures and Buildings*, 1998, 128(4), 351–358.
[16] Smith S. D., Modelling and experimentation of the concrete supply and delivery process, *Civ. Eng. Env. Syst.*, 1999, 16(2), 93–114.
[17] Abourizk S. M., Halpin D. W., Wilson J. R., Visual interactive fitting of beta distributions, *J. Constr. Engrg Mgmt, ASCE*, 1991, 117(4), 589–605.
[18] Law A. M., Kelton W. D., Simulation Modeling and Analysis, Volume 3, McGraw-Hill, Inc, New York, N.Y., U.S.A., 2000.
[19] Devroye L., Non-Uniform Random Variate Generation, Springer-Verlag, New York, N.Y., U.S.A., 1986.
[20] Cheng R. C. H., The generation of gamma variables with non-integral shape parameter, *Appl. Statist.*, 1977, 26, 71–75.
[21] Shanks R., Abelson R., Scripts, Plans, Goals and Understanding, Lawrence Erlbaum Assoc., Hillsdale, N.J., U.S.A., 1977.
[22] Kolodner J., Case-Based Reasoning, Morgan Kaufmann, San Mateo, California, U.S.A., 1993.
[23] Tah J. H. M., Carr V., Howes R., Information modeling for case-based construction planning of highway bridge projects, *Adva Engrg Software*, 1999, 30, 495–509.
[24] Kumar H., Krishnamoorthy C., A framework for case-based reasoning in engineering design, *Artif. Intell. For Engrg Design, Analysis, and Manufact.*, 1995, 9, 161–182.
[25] Reich Y., Fenves S. J., System that learns to design cable-stayed bridges, *J. Infrastruct. Syst.*, 1995, 121(7), 1090–1100.
[26] Dzeng R. J., Tommelein I. D., Boiler erection scheduling using product models and case-based reasoning, *J. Constr. Engrg Mgmt, ASCE*, 1997, 123(3), 338–347.
[27] Morcous H., Rivard A. M., Hanna A. M., Case-based reasoning system for modeling infrastructure deterioration, *J. Comp. Civil Engrg, ASCE*, 2002, 16(2), 104–114.
[28] Arditi D., Tokdemir O. B., Comparison of case-based reasoning and artificial neural networks, *J. Comp. Civil. Engrg, ASCE*, 1999, 13(3), 162–169.
[29] Aamodt A., Plaza E., Case-based reasoning: Foundational issues, methodological variations, and system approaches, *AI Communications*, 1994, 7(i), 39–59.
[30] Kriegsman M., Barletta R., Building a case-based help desk application, *IEEE Intell. Syst.*, 1993, 8(6), 18–26.
[31] Neale R., Ameen J., Smith S. D., Discussion on earthmoving productivity estimation using linear regression techniques, *J. Constr. Engrg Mgmt., ASCE*, 2001, 127(1), 88–89.
[32] Quinlan J. R., Introduction to decision trees, *Machine Learning*, 1(1), 81–106.
[33] Hand D., Mannila H., Smyth P., Principles of Data Mining, Massachusetts Institute of Technology, Cambridge, Mass., U.S.A., 2001.
[34] Tokdemir O. B., Predicting the outcome of construction litigation using case based reasoning, MS Thesis, Illinois Institute of Technology, Chicago, U.S.A., 1997.
[35] Watson I., Applying Case-Based Reasoning: Techniques for Enterprise Systems, Morgan Kaufmann, San Mateo, Calif., U.S.A., 1997.
[36] Naylor T. H., Computer Simulation Experiments with Models of Economic Systems, John Wiley and Sons, New York, N.Y., U.S.A., 1971.
[37] Smith S. D., Osborne J. R., Forde M. C., Analysis of earth-moving systems using discrete-event simulation, *J. Constr. Engrg Mgmt, ASCE*, 1995, 121(4), 388–396.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Advanced Engineering Informatics 18 (2004) 17–28
ADVANCED ENGINEERING INFORMATICS
www.elsevier.com/locate/aei

Estimating the productivity of cyclic construction operations using case-based reasoning

Darren Graham, Simon D. Smith*

School of Engineering and Electronics, The University of Edinburgh, Alexander Graham Bell Building, Edinburgh EH9 3JL, UK

Received 5 March 2004; revised 5 March 2004; accepted 15 March 2004

Abstract

The estimation of productivity rates in cyclic construction processes is a difficult, but essential task in the planning of construction projects. The conventional method—a calculation by intuition—for devising such estimates is often laborious and inaccurate. To overcome these difficulties this paper proposes that Case-Based Reasoning could be employed to provide estimates from historical cases of construction. To this end a model, Case-Based Estimator (CBE) is presented. CBE is a small model, consisting of five input features, one output and a small case base, and it is suggested that the findings within this paper could be applied to any small model of similar characteristics—not only within the construction field. The development and implementation of the model is discussed and an experiment was performed to assess the ability of two retrieval mechanisms (one a simple mathematical formula, the other an adaptation of the ID3 decision tree generating algorithm) to measure similarity. The simple formula was found to be more preferable, both in terms of consistency and development effort.

CBE was validated, not only against the performance of past operations (which were not used in the model development), but also against estimates provided by a professional construction planner. The model was found to provide more precise and consistent estimates than the planner, with 90% of the estimates being within a 10% relative error of the observed value. However, CBE did not perform as accurately when estimating operations which were thought to occur only rarely (outliers).

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Cyclic construction operations; Case-based reasoning; Productivity rates; Estimation

1. Introduction

Many engineering and construction projects are based upon repetitive or cyclical processes, such as concreting or earthmoving. Successful planning of these processes can result in considerable amounts of time and money being saved over the duration of a project; these savings translate to increased profits for companies, enabling the companies to compete more successfully and enable their clients' expectations to be met more consistently. The planning of cyclic construction processes is often difficult due to their stochastic nature and in order to reduce this difficulty, many planning engineers maintain a large databank of productivity rates recorded from past projects. For a new construction operation, these productivity rates are adjusted, taking into account specific site factors and conditions that may influence productivity [2]. There are two potential problems with this method: firstly, the databank of productivity rates is usually stored in paper form [9], which makes the identification of appropriate productivity rates difficult and laborious; and, secondly, a recent investigation found differences between the average productivity rates actually measured on site and those used by planning engineers [2]. To reduce the effort required to plan construction operations, and to improve the accuracy of productivity estimates this study proposes the use of an Artificial Intelligence technique, Case-Based Reasoning (CBR). CBR looks for previous cases that are similar to the current problem and reuses them to solve the problem [10]. It is very applicable in domains where there are large amounts of historic information available, such as in the planning of construction operations.

To demonstrate the potential ability of CBR to estimate cyclic construction productivities, this paper presents a case-study of the application of a CBR model, Case-Based Estimator (CBE), to concreting operations. CBE has been

* Corresponding author. Tel.: +44-131-650-7159; fax: +44-131-650-7781.
E-mail address: simon.smith@ed.ac.uk (S.D. Smith).

1474-0346/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aei.2004.03.001

developed in Microsoft Excel, using VBA macros to automate the CBR process.

The structure of this paper is as follows:

- Literature review
- Overview of concreting operations
- Summary of the collection of data
- Development of CBE
- CBE from the user's point of view
- Validation of CBE
- Experimentation with the model set-up
- Discussion and conclusions

2. Literature survey and review

CBR applications are becoming popular in the fields of construction management and civil engineering [19]. In the area of design there have been a number of CBR models developed, including: CASETOOL, a system for bridge design [12] and BRIDGER, a system to aid the conceptual design of cable-stayed bridges [15]. Faltings et al. [4] and Shih [17] applied CBR to architectural design, and Flemming [6] and Rivard et al. [16] addressed building design issues. In the area of planning, Tah et al. [19] created CBRidge Planner to aid the planning of highway bridge projects. Another planning model, CasePlan was developed for scheduling and planning boiler erection by Dzeng and Tommelein [3]. All of these CBR models and systems design or plan but do not predict. In the more recent past, CBR has been used successfully in prediction and estimation problems [13]. Some examples of the few applications of CBR to prediction and estimation are Arditi and Tokdemir [1] to predict the outcome of construction litigation; Yau and Yang [21] developed a model to estimate construction duration and costs of building construction projects; and a model to estimate productivity rates of masonry block wall construction was developed by Karshenas and Tse [9].

This study intends to use CBR to estimate the output of cyclical construction processes and as such shall provide a unique application of CBR in the construction management field. A further aim of this study is to allow the methods used in this specific application to be applied to other prediction and estimation problems. By comparing the effect of different retrieval processes it is hoped that this study will reach beyond the field of construction management.

3. Concreting operations

'Concreting operations' is a general term used to describe the process of concrete supply, delivery and placement, which is shown in Fig. 1. This process begins at a concrete batching plant where a truck-mixer (truck) is batched with a stipulated volume of fresh concrete. The truck subsequently departs to its destined construction site. On arrival, it will enter 'service' (if no other trucks are present) or join the back of the queue of waiting trucks. Service requires the maneuvering of the truck into position and discharging its load into the hopper of the pump. The pump relays the concrete into prepared framework. The empty truck is then washed out at a pre-assembled area. When clean, the truck returns to the batching plant, to receive another load of concrete, therefore making the process cyclical. Finally, the concreting operation is completed when a specified amount of concrete has been placed.

Concreting operations are stochastic and the events that occur within the system (e.g. truck arrival times and pump start times) take place at irregular intervals [18]. An under-utilization of the plant and labor is to be expected, due to the queuing of trucks or idleness of the pump, as it is improbable that a truck arrives simultaneous to the previous truck departing. This queuing (or idleness) results in

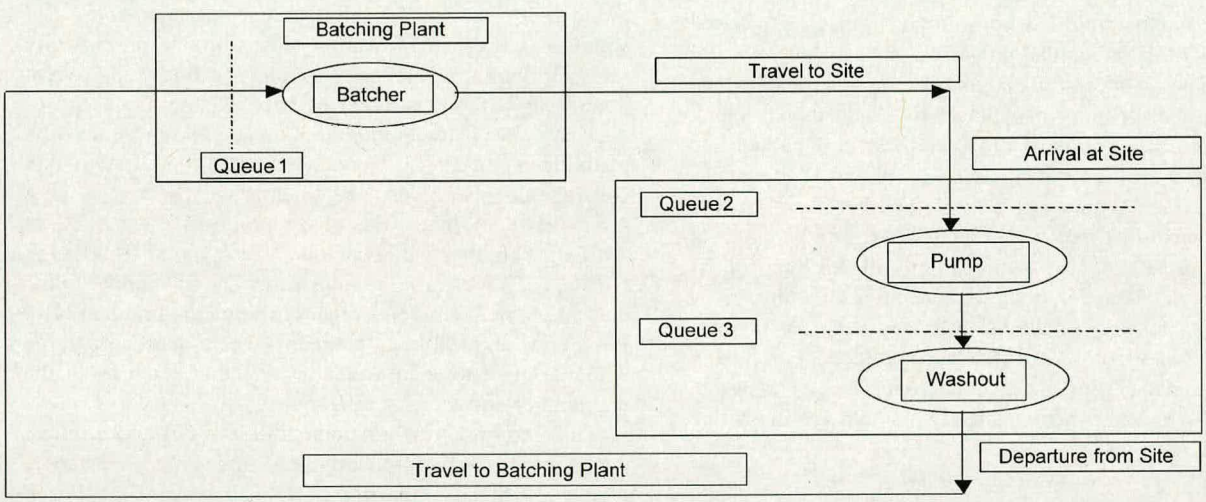


Fig. 1. Schematic diagram of the concrete supply, delivery and placement process.

ariability in the overall productivity of the operation. Having explained the process of concrete supply, delivery nd placement, the steps involved in the development of a BR model can now be stated as:

- Gathering of real construction site data
- Development of CBR model, *Case Based Estimator (CBE).*
- Using CBE
- Validation of CBE
- Experimentation with the model set-up

. Gathering of real construction site data

The data on which CBE is based was collected during a ime study of four construction projects in Scotland, UK. hese projects involved constructing wastewater treatment lants. In total, 220 concreting operations were observed and he following variables were recorded: pour date and type, veather, batch time, arrival time on site, position of truck ime, start and completion of pumping times, washout time, arget and actual slump, batch quantity, pour volume, truck vait time (at batcher and pump), batcher and pump idle times nd the number of trucks involved in the operation.

The data that was used to validate CBE was collected rom a construction project, differing in type and location rom those detailed above. This project took place in the North of England, UK, and involved the construction of a 5 ane highway viaduct. From the data set collected, 20 perations were selected for use in the validation process.

5. Development of a CBR Model, CBE

CBR is a cyclical process, which has four main aspects 10]. They are:

a) Case representation and storage: organization of the information describing the problems and their solutions.
b) Case retrieval, which searches the case library for the case(s) that best match the problem (target case).
c) Case adaptation, which revises the retrieved case(s) to fit the current problem context.

Case accumulation, which is responsible for the storage f new cases and the update of existing cases.

These aspects have been incorporated in the development f CBE and the following is a description of each stage of he development process.

5.1. Case representation and storage

Watson [20] defined a case as a contextualized piece of nowledge representing an experience. It contains the past lesson that is the content of the case and the context in which the lesson can be used. A case can be an account of an event, a story, or some record—in the current study it is the productivity of a particular concreting operation—and typically comprises:

(i) The problem that describes the state of the world when the case occurred. In CBE the problem is defined by five input features (attributes), they are: number of truck-mixers involved in an operation; pour volume, pour type, month of pour and weather. Each feature is separated into numerical ranges, which are indexed (Fig. 2). There are undoubtedly more features, which influence concreting operations than those considered in this study, however, if an accurate and robust model is produced then this simplicity is justified [9].
(ii) The solution that states the derived solution to that problem. In CBE the solution contains output features which were measured on construction sites, such as: operation completion time, truck idle time, and pump idle time. Additional output features that have been derived from these measured values are: productivity, truck and pump utilization rates.
(iii) A robust case base, containing a representative and well-distributed set of cases, is the foundation for a good CBR system [11]. At the beginning of the study the case base in CBE contained 220 records, formed from the data gathered from the construction sites mentioned previously. The size of the case base is dynamic: any duplicate cases must be removed before CBE can be used and the case base stores the solution as a new case. The next stage in the process is case retrieval.

5.2. Case retrieval

A CBR system derives its power from its ability to retrieve relevant cases quickly and accurately from its case base [1]. Considering the need for efficiency, the retrieval process in CBE consists of the following process:

1. Check if a past solution matches the current target case. If so, the matching solution is presented to the user for consideration.
2. If there is no matching prior solution, the retrieval mechanism is invoked. The retrieval mechanism consists of four stages:

1. *The measurement of the similarity between a new case and each case contained in the case base.* This is undertaken using two mathematical formulae: feature counting (see Appendix A), where all the attributes (or features) are given an equal importance in the calculation of similarity; and weighted feature counting (see Appendix A), where attributes are given weightings. In both formulae, each attribute of the new case is compared with the corresponding

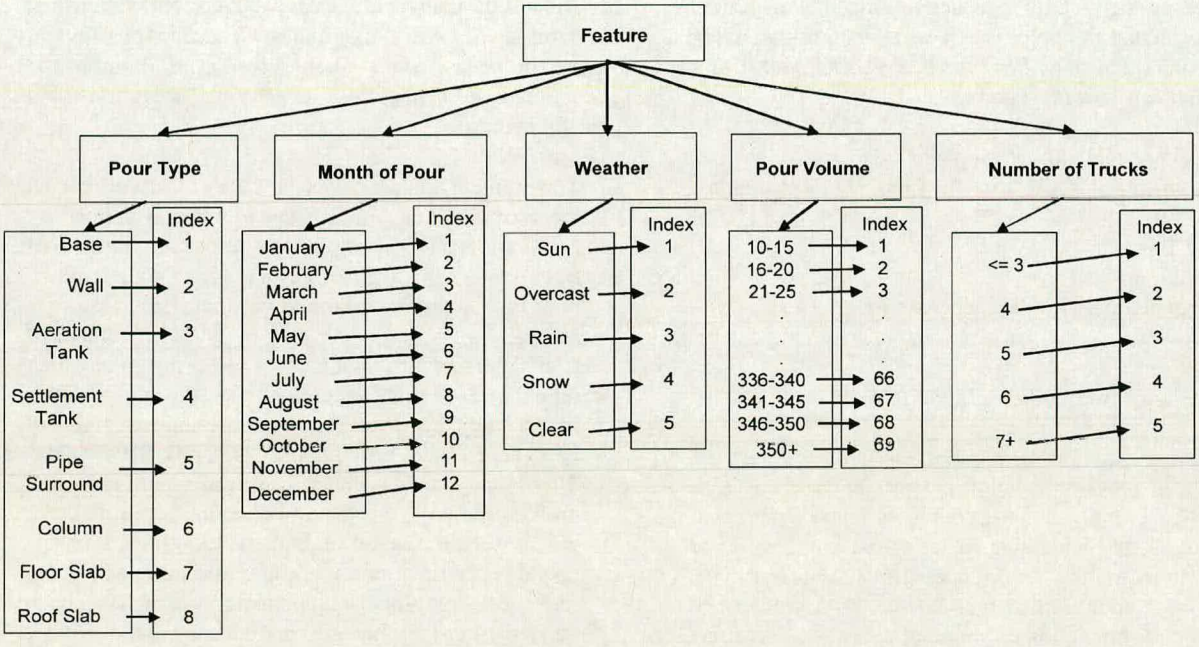


Fig. 2. Hierarchy of case representation in case-based estimator.

attribute of each case in the case base (e.g. month of pour compared with month of pour) and if the attributes match, a value of 1 is entered into the formula for that particular attribute. This continues until all the attributes contained in a case have been examined. It is at this point that an average of the results is taken which is taken as the *similarity measurement* (in percentage form). To illustrate further the following example is presented:

The attributes of the new case are: number of truck-mixers involved in the operation = 3; pour volume = 56 m^3; pour type = base pour; month of pour = July; and, weather = sunny. These values have been indexed using the method shown in Fig. 2, and are summarised in Table 1. For the purposes of this example this new case shall only be compared with the first case contained within the case base. The indexed attributes of this case are: number of truck-mixers involved in the operation = 5; pour volume = 61; pour type = 1; month of pour = 7; and, weather = 5. In this instance, only the pour type and the month of pour indices match. This results in a similarity between the two cases equals to:

$$\frac{0(\text{no.trucks}) + 0(\text{pour volume}) + 1(\text{pour type}) + 1(\text{month of pour}) + 0(\text{weather})}{5} \times 100 = 40\%$$

The example was simplified as only the feature counting measurement was used to calculate the similarity. In CBE, the results of feature counting are used to provide a 'rough guide' as to the most similar cases. The weighted version is included to allow more significance to be allocated to certain attributes, thus making the model more realistic, and this weighted measurement forms the basis for judging similarity. The increased realism that is attached to the weighted measure comes at a cost, unfortunately, due to the difficulties involved in accurately calculating the weightings of each attribute. To overcome these difficulties a decision tree can be used to calculate the weightings.

The manual creation of a decision tree, even for a data set of only 220 cases would be extremely cumbersome, and therefore a method of automating the formation of a decision tree was required. The automating method used in CBE, is the ID3 algorithm [14], which calculates the weighting of an attribute using a metric known as the information gain. This metric relies on the concept of entropy, which is a measure of the uncertainty in a probability distribution. In CBE, the output of concern is productivity, which is not separated into discrete categories, but is an actual number. Assuming that no two productivities in the case base are identical the probability distribution will have 220 separate productivities, all with an equal chance of occurring. This situation results in an extremely complex calculation of entropy. To reduce the complexities in calculating entropy, the productivity of each case in the case base is compared with the average of all the cases in

able 1
ndexed values for similarity measurement example

eature	Actual information	Index number
our type	Base	1
Ionth of pour	July	7
umber of trucks	3	3
our volume (m^3)	56	10
Veather	Sunny	1

able 2
eature weightings calculated for the original case base

eature	Weighting
ype of pour	0.20
Ionth of pour	0.09
umber of trucks	0.29
our volume	0.34
Veather	0.08
otal	1.00

the case base. Each case is then indicated as having productivity above or below the average of the case base. Thus the complexity of the situation has been substantially reduced (from 220 possibilities to 2) as the probability distribution only consists of the chances that the productivity of a case is above or below the average. The information gain is used as the weighting of each attribute in the subsequent calculation of similarity between a new case and each case in the case base.

The weights calculated for the initial case base, using the ID3 algorithm are shown in Table 2. The weightings are as expected, with the most important features being the number of trucks and the pour volume. However, as the weightings are key to an accurate retrieval process it is important to confirm that they are set at an appropriate level. A sensitivity analysis was performed to determine the effect on the retrieval process, of changing a feature weight by -50, -20, -10, $+10$, $+20$, and $+50\%$ (while the other features remain constant). Table 3 shows the index numbers of the five cases which were most similar to the original

able 3
op-five ranked cases in weightings tests

eature	Ranking	Percentage difference from original weighting						
		−50	−20	−10	0	10	20	50
our type	1	9	9	9	9	9	9	181
	2	181	181	181	181	181	181	9
	3	178	17	17	17	17	17	17
	4	17	178	11	11	11	11	11
	5	14	11	14	14	14	14	14
Ionth of pour	1	181	9	9	9	9	9	9
	2	9	181	181	181	181	181	181
	3	17	17	17	17	17	17	17
	4	11	11	11	11	11	11	11
	5	14	14	14	14	14	14	14
Io. trucks	1	9	9	9	9	9	9	181
	2	181	181	181	181	181	181	9
	3	14	17	17	17	17	17	17
	4	17	14	11	11	11	11	178
	5	11	11	14	14	178	178	11
Io. of trucks	1	9	9	9	9	9	181	181
	2	181	181	181	181	181	181	9
	3	14	17	17	17	17	17	17
	4	17	14	11	11	11	11	178
	5	11	11	14	14	178	178	11
our volume	1	9	9	9	9	9	181	181
	2	181	181	181	181	181	9	9
	3	14	14	17	17	17	17	17
	4	13	17	14	11	11	11	11
	5	11	11	11	14	178	178	178
Veather	1	9	9	9	9	9	9	9
	2	181	181	181	181	181	181	181
	3	17	17	17	17	17	17	14
	4	14	11	11	11	11	11	17
	5	11	14	14	14	14	14	11

target case (0 percentage difference from the original weighting), and the consequences—in terms of the five most similar cases—of adjusting the weights by the amounts mentioned above for each feature in turn. From the results it can be concluded that small changes (±20%) in the weighting of a feature had negligible effect on the case selected. However, a change of ±50% resulted in a different case being suggested by CBE for: pour type, month of pour, number of trucks, and pour volume.

To investigate the effect of this change, the productivities of the observed operation were compared with those suggested by the two proposed cases (Fig. 3). Clearly, the difference between productivities is small, and it can be concluded that the feature weightings are set at an appropriate level.

2. *Sorting of the results.* The cases which are below the specified threshold of similarity are removed from the table of results. The remainder are sorted into descending order of rankings and the top ten most similar cases are presented in the final results table.

User selection. The user selects the case to be used to provide an estimate of a new concreting operation. This case is usually the top ranking case in the final results table, however, the user is free to choose another based on intuition.

Retrieval of the case information. The relevant information contained in the selected case e.g. the productivity is retrieved from the case base for use in the remainder of the CBR cycle.

At this stage of the study, the accuracy of the retrieval mechanism must be verified before continuing to develop the model. Gonzalez et al. [8] presented a method to verify the retrieval mechanism of a CBR system, known as the Retrieval Test. This uses each case held in the case base as a target case. In order for a test to be marked as successfully executed, the historical case that matched the target case should be found as top-ranked in the results report. The test

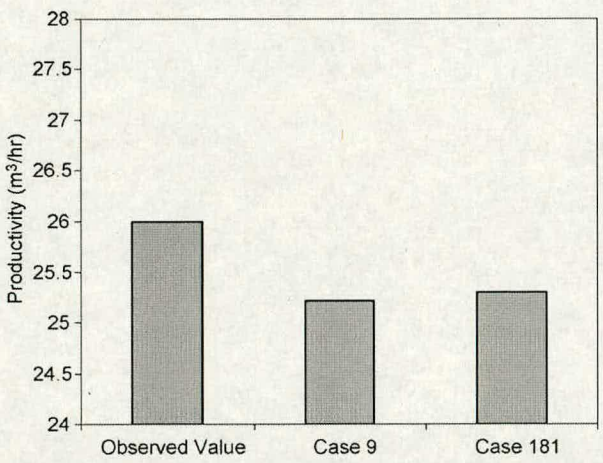


Fig. 3. Plot of productivities estimated by the top-ranked cases in the weightings test against the observed productivity.

Table 4
Results from the retrieval test

Retrieval test no.	No. cases retrieved	Case base size	Percentage of cases retrieved
1	178	220	81
2	181	181	100

was undertaken for the original case base comprising 220 cases and the results are shown in Table 4. This test highlighted a number of duplicate cases, which confused the retrieval mechanism, resulting in only 81% of cases being retrieved correctly. The duplicate cases were removed, leaving a case base of 181 cases. The Retrieval Test was undertaken again, and this time the results show that the retrieval mechanism identified 100% of the cases (Table 4). For the purposes of this study, the mechanism can be considered verified.

5.3. Case adaptation

Case adaptation is the process of adapting the solution stored in a retrieved case to the needs of the current (target) case. While adaptation is useful in many situations, it is by no means essential [20]. CBE uses an adaptation method known as Null Adaptation [20], which involves forming a new case comprising the features of the target case and the solution (productivity estimates) from the selected case. It is necessary to verify that this adaptation method is estimating concreting productivity accurately. This verification task shall be undertaken using the Adaptation Test developed by Gonzalez et al. [8]. Before the test methodology can be explained, some terms used in the Adaptation Test should be explained:

- Resulting Acceptability Criteria (RAC), this serves to determine whether an individual test case has been solved correctly by the CBR system. In CBE, this is a measure of the estimated productivity compared with the observed productivity, for each case. For the purposes of this study, the RAC level has been set at 80%.
- Relative Error (RE), this is the percentage difference between the productivity estimated by CBE and the observed productivity, for each case.
- System Validity Criteria (SVC), serves to determine whether, in light of the executed suite of test cases, the system can be considered valid, or at this stage, verified. Due to the small size of the case base, the SVC level has been set at 80%. As the case base grows different situations shall be recorded, therefore making CBE more knowledgeable about concreting operations, and possibly allowing an increase in the SVC level.

The Adaptation Test consists of removing a case from the case base and using it as a target case. Although the case is

able 5
esults from the adaptation test

ases with AC 80% or above	Case base size	SVC (%)	Average RE (%)
63	181	90.06	8.39

o longer contained in the case base, CBE retrieves the most imilar case(s) from the case base and presents them to the ser. Since the retrieval mechanism has already been erified, this test isolates and evaluates the Null Adaptation rocess of CBE. The results of the Adaptation Test (Table 5) how that the cases suggested by the system in 163 cases out f 181 cases were above the RAC value of 80%, resulting in n SVC of 90%. Therefore, the adaptation process of the BE model can be considered verified. The next stage in the BR cycle is case accumulation.

.4. Case accumulation

Case accumulation is the final stage of the CBR process nd involves updating the case base with the target problem nd its solution. Fenves et al. [5] stated that the more cases ontained in the case base, the better the performance of the system. Therefore, the case accumulation process provides a CBR system with its learning capacity. After a new case is updated, the ID3 algorithm is invoked to calculate new weightings for the features of the system.

6. Using CBE

To obtain productivity estimates from CBE, the user must input the attributes of a particular construction operation into the system's data-entry screen via a series of drop-down boxes (Fig. 4). On receipt of this information, the cyclical CBR process is invoked and a report containing the top-ten ranked solutions, as measured using both the feature counting and weighted feature counting methods, are presented to the user (Fig. 5). The user has a number of options at this stage: to select a case; to reject all cases; or, to view the productivity information held in the ten cases, and subsequently select a case or reject all cases. If the user selects a case, the productivity information from that case is combined with the attributes of the target case, forming a new case. Finally, the user is presented with the new case and the productivity estimates can be used in the planning of concreting operations.

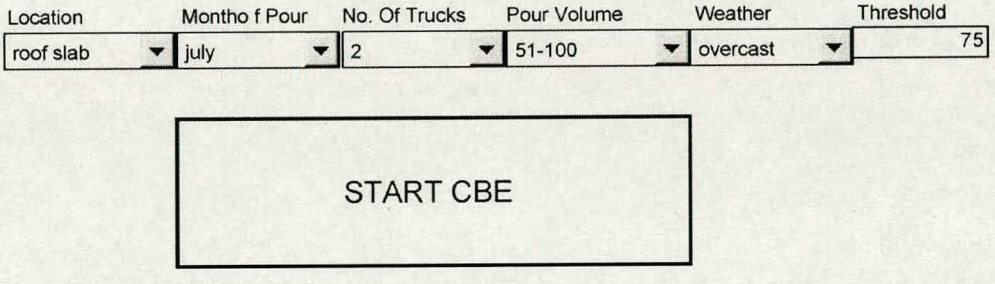


Fig. 4. CBE user input screen.

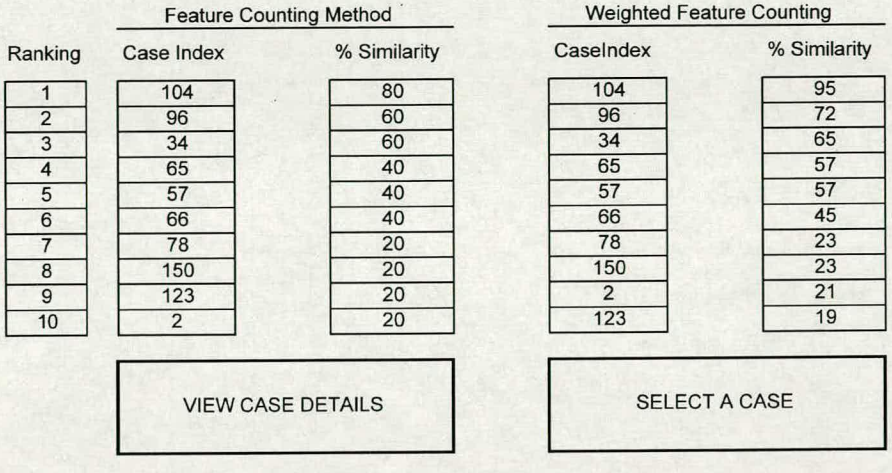


Fig. 5. CBE retrieval results report.

7. Validation of CBE

The concreting operations model is intended to represent the existing system, not replace it, to allow estimates of the system's output to be produced. Therefore, the output of CBE can be compared with observed values from the real system to determine if it is valid—that is to say, the model has the ability to predict the behavior of the real systems variables, irrespective of whether or not the assumptions the model is based upon, are reasonable [7]. As mentioned earlier, concreting operations on an independent construction project were observed, and 20 pours were selected for use in this validation test. The features of each case were used as a target case, and the output of the CBE model (estimated productivity) was compared to the observed productivity values using the techniques mentioned in the Adaptation Test. As before, the RAC and SVC levels were set at 80%. The results of the validation test are shown in Table 6 and 90% of the validation observations had a RE of less than 10%. A plot of the observed productivity against the estimated productivity (Fig. 6) further indicates the ability of the CBE model to provide close estimates of concreting operations. The results are sufficiently accurate to tentatively declare CBE to be a valid model.

The data used in the original validation test discussed above was from a source that was different in nature

Table 6
Results from the validation study

Validation observation	Data source	Pour type	Month of pour	No. of trucks	Pour volume (m^3)	Weather	Productivity (m^3/h)	RE (%)	% age cases within 20% RE	% age cases within 10% RE
1	Observed CBE	Wall	May	3	53	Sunny	13.82	0.07		
							13.83	0.07		
2	Observed CBE	Base	May	4	128	Sunny	24.3	8.77		
							22.17			
3	Observed CBE	Wall	May	3	56.5	Overcast	13.89	9.58		
							12.56			
4	Observed CBE	Wall	May	3	64	Overcast	10.94	3.38		
							11.31			
5	Observed CBE	Base	May	8	144	Sunny	21.82	0.00		
							21.82			
6	Observed CBE	Wall	May	2	56	Sunny	10.57	8.89		
							11.51			
7	Observed CBE	Wall	May	3	56	Sunny	15.85	21.39		
							12.46			
8	Observed CBE	Wall	May	4	55	Clear	10.89	3.03		
							11.22			
9	Observed CBE	Wall	May	2	53	Sunny	11.65	0.52		
							11.59			
10	Observed CBE	Wall	June	4	63	Overcast	11.42	1.66		
							11.61			
11	Observed CBE	Wall	June	3	59.5	Sunny	8.73	4.35		
							9.11			
12	Observed CBE	Wall	June	4	64	Overcast	13.33	6.90		
							12.41			
13	Observed CBE	Floor slab	June	2	48	Sunny	12.26	5.14		
							11.63			
14	Observed CBE	Wall	June	2	51	Sunny	13.72	7.94		
							12.63			
15	Observed CBE	Base	September	3	65	Sunny	13.27	2.79		
							13.64			
16	Observed CBE	Wall	September	4	51	Sunny	15.38	0.07		
							15.39			
17	Observed CBE	Base	September	4	66.5	Overcast	14.25	4.28		
							13.64			
18	Observed CBE	Wall	September	3	42.5	Rain	11.33	9.53		
							12.41			
19	Observed CBE	Roof slab	October	3	80	Overcast	18.82	21.63		
							14.75			
20	Observed CBE	Base	February	7	180	–	22.64	6.71		
							24.16	6.71		
									90	90
							Average RE	6.33		

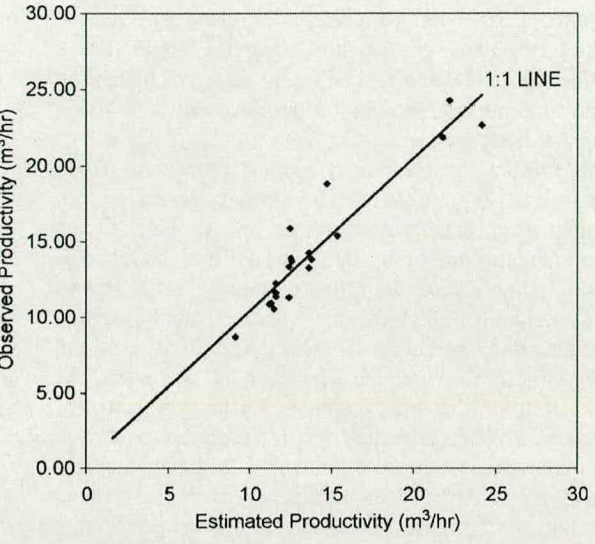


Fig. 6. Observed productivity plotted against estimated productivity for CBE.

and location to the data sources which comprise the case base. Whilst this provides a useful test of CBE's ability to generalize, further testing of this ability is important. An further validation test has therefore been carried out to provide more evidence of the capacity of the model to estimate cases which are outliers to the cases contained in the case base. Such cases are indicated below, with the reason for being an outlier in brackets:

- 5 (the large number of trucks),
- 13 (the type of pour),
- 19 (the type of pour), and
- 20 (the month of pour).

Table 7 shows a comparison of the average RE produced by CBE for the four outlier cases with the average RE produced by CBE for the remaining non-outlier cases. Unsurprisingly, due to the nature of the model (it recreates history) the results show that CBE produces better results (roughly 5% less RE) when estimating the cases which have been deemed to be non-outliers. The results are still promising, and clearly the more knowledge the model acquires (more cases), the greater the accuracy and reliability when estimating new cases.

Table 7
Validating the model for outliers

Case type	Case numbers	Average RE(%)
Outlier cases	5,13,19,20	11.2
Non-outlier cases	1,2,3,4,6,7,8, 9,10,11,12,14, 15,16,17,18	5.82

8. Experimentation with the Model Set-up

The basic CBE model has now been developed and validated to a degree, and there is now the potential to investigate methods of improving each stage of the CBR cycle within the model. One such investigation that has been undertaken is a method to enhance the mechanism which calculates the similarity between cases and retrieves the appropriate cases from the case base. The method in question is an adaptation of the ID3 algorithm, which was used successfully to generate weightings for the attributes in the case base.

In this instance the ID3 algorithm shall be utilized to calculate the similarity between a new case and each case in the case base. This version of the ID3 algorithm is again based on a calculation of entropy, which is simple in this instance (as the probability distribution is based on the chance that the attribute of the new case is either equal to or not equal to the attribute of the case held in the case base). A calculation of the information gain metric for a case is performed using the combined entropy values of each attribute in that particular case. The measurement of similarity is simply the ratio of information gain from each case in the case base to the information gain of the whole case base. A percentage value of similarity is attached to each case in the case base, and steps (2)–(4) of case retrieval (see above) are performed.

To compare the performance of applying the ID3 algorithm in this manner with the original retrieval method, the validation data was used as target cases and the estimates that were provided by the most similar case in each instance (from ID3 and original retrieval mechanism) were compared with the observed values (see Fig. 7). Fig. 7 shows that similar results are achieved using both retrieval methods, but the original retrieval method performed more consistently.

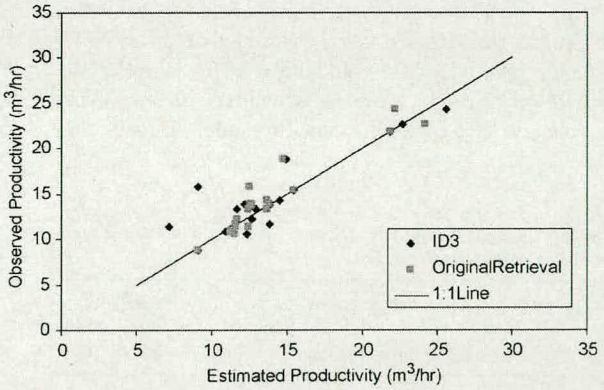


Fig. 7. Observed productivity plotted against the estimated productivity from the ID3 algorithm.

9. Comparison between CBE and conventional estimating

A conventional method used to estimate the productivity of cyclic construction operations is to adjust the productivity rates which are maintained in a data bank. Dunlop and Smith provided the details of three productivity rates—one each for base, wall and column pours—obtained from a major UK contractor's planning office [2]. For pours based on a mobile pump, whose placing capacity is 20–30 m^3/h, the following placing speeds are used in the planning of concreting operations:

- Base Pours 13.3 m^3/h
- Wall Pours 9.6 m^3/h

These estimations have not been adjusted specifically for each operation, as would happen in a real situation. However, it may be useful to calculate the average of the estimates provided by CBE over the whole validation data set (for base and wall pours), and to compare these averaged results with the 'averaged' planning estimates mentioned above. Table 8 shows this comparison and it can be seen that the productivity estimates provided by CBE reflect the observed results more closely than the estimates obtained in the conventional manner. This is to be expected as the estimates from the planner have not been adjusted from the average to take in to account each specific situation, and planners historically predict conservative estimates to allow for the uncertainties in the construction processes. Nevertheless, the results from CBE are extremely close to reality, and as such the model has the potential to allow planning engineers to organize, with a high degree of confidence, the resources required to undertake the construction operations.

10. Discussion and conclusions

The conventional methods used by planning engineers in estimating construction productivity rates are laborious and can be inaccurate. This paper proposed that an artificial intelligence method, CBR could be used to simplify the determination of productivity rates, whilst also increasing their accuracy. This paper presented a model of concreting operations, CBE. CBE was developed using information collected from UK construction projects. Originally 18 variables were collected from each project, but they were reduced in number to five in the development of CBE, to ensure simplicity.

Table 8
A comparison between the average estimates produced by CBE and those obtained using conventional methods

Pour type	Actual productivity rates (m^3/h)	Productivity estimates using formation from a professional planner m^3/h)	Productivity estimates from CBE (m^3/h)
Base	19.26	13.33	19.09
Wall	12.42	9.6	12.16

CBE follows the traditional cyclical process of CBR, which is: case representation and storage; case retrieval; case adaptation; and case accumulation. All parts of the CBR process are important, but it is vital that the retrieval mechanism is accurate and reliable. Initially, in CBE, two simple mathematical formulae (feature and weighted feature counting) were used to assess the similarity between a new (target) case and the stored cases. To assess the effectiveness of these measures a verification test was undertaken. This test consisted of using each case contained in the case base, in turn, as a target case and checking that the case suggested by both measures was the case being tested. Initially the retrieval mechanism was not performing with perfect accuracy, and this was found to be due to the case base containing duplicate cases. The verification test proved to be an effective method of uncovering all the duplicate cases contained in the case base, and after the removal of these cases the retrieval mechanism worked with 100% accuracy.

CBE is intended to represent the concreting system as closely as possible, in order to allow estimations of the productivity to be produced. To determine if CBE is valid, the productivity information from 20 concrete operations observed on an independent construction project was tested. CBE achieved productivity estimates to within 10% accuracy, in 90% of the validation test pours, and can tentatively be declared valid.

The ability of the model to provide good estimates when a case was declared to be an outlier (rare event) was assessed. It was found that CBE cannot produce the same level of accuracy in outlying cases, but intuitively this could be expected. CBR works by using the knowledge contained within the model, therefore a lack of cases which have knowledge about each feature will produce a weaker model. One method of overcoming this drawback to CBR is through building and maintaining an extensive case base and this is part of the future plan with regards to CBE.

To calculate the weightings of each feature (for use in the weighted feature counting similarity measurement) the ID3 algorithm is used. This algorithm invokes the formation of a decision tree which is used to assess the relative importance of each feature. In CBE, the ID3 algorithm is found to be an acceptable method of calculating weightings; the weightings of each feature were as predicted from the researchers knowledge, and a sensitivity analysis of the weightings found that even a significant change in the weightings did not produce a large difference in the estimate of productivity produced by CBE. Although it is recommended here that the ID3 algorithm should be used to calculate weightings, the amount of time required for development and implementation was substantial when compared with

ne remainder of the retrieval process. In an attempt to educe the development time of any future models, and to nprove the retrieval mechanism, the ID3 algorithm (used o calculate the weightings of features) was adapted so that it ould provide the weightings of each stored case in relation o a target case, which is effectively a similarity measure- ient. The ID3 algorithm was found to provide similar, but ess consistent results to the original retrieval method. Although the development time of the ID3 algorithm was ess than the original retrieval method, it is worth emembering the importance of the retrieval mechanism n the CBR process. Therefore, it was decided that the riginal method of retrieval should be used in the remainder f the study.

- If, month of pour index (target case) = month of pour index (stored case), month of pour = 1, otherwise month of pour = 0.
- If, number of trucks index (target case) = number of trucks index (stored case), number of trucks = 1, otherwise number of trucks = 0.
- If, pour volume index (target case) = pour volume index (stored case), pour volume = 1, otherwise pour volume = 0.
- If, weather index (target case) = weather index (stored case), weather = 1, otherwise weather = 0.

The above values are then used in the calculation of similarity. For feature counting the formula is:

$$\text{imilarity (\%)} = \frac{(\text{pour type} + \text{month of pour} + \text{number of trucks} + \text{pour volume} + \text{weather}) \times 100}{5}$$

A comparison between the estimates produced by CBE nd a professional construction planner was made, and CBE vas found to have greater accuracy in predicting the roductivity of operations. It may be concluded that CBE is seful in planning and estimating concreting operations as it an reduce the data intensive methods currently used by lanning engineers and it produces close estimates of oncreting operations productivity rates.

CBE in its present form is a small model. As such, the findngs of this study could plausibly be applied to any small model vith a simple set of inputs and outputs and a small case base.

Some suggestions for future work that may be carried out n this area are:

- Increase in the complexity of the model by increasing the number of variables and the number of cases stored.
- Further validation of CBE.
- Further validation of the ID3 algorithm method of selecting the best case to represent a target case.
- Development of a model of a similar cyclical construction process, such as earthmoving, to determine the applicability of CBR in estimating all such processes.

Appendix A

The key to feature counting is the question are the eatures (indexed) of the target case equal to the features indexed) of the stored case? To answer this question the ollowing are calculated:

- If, pour type index (target case) = pour type index (stored case), pour type = 1, otherwise pour type = 0.

For weighted feature counting the formula for similarity equals:

$$a = \omega_{\text{pour type}} \times \text{pour type}$$

$$b = \omega_{\text{month of pour}} \times \text{month of pour}$$

$$c = \omega_{\text{number of trucks}} \times \text{number of trucks}$$

$$d = \omega_{\text{pour volume}} \times \text{pour volume}$$

$$e = \omega_{\text{weather}} \times \text{weather}$$

$$\text{Similarity (\%)} = \frac{(a+b+c+d+e) \times 100}{5}$$

ω represents the weightings of a particular attribute, as calculated using the ID3 algorithm.

References

[1] Arditi D, Tokdemir OB. Comparison of case-base reasoning and artificial neural networks. J Comp Civil Engng 1999;13(3): 162–9.
[2] Dunlop P, Smith SD. Estimating key characteristics of the concrete delivery and placement process using linear regression analysis. Civil Engng Environ Syst 2003;20(4):273–90.
[3] Dzeng R-J, Tommelein ID. Boiler erection scheduling using product models and case-based reasoning. J Construct Engng Mgmt 1997; 123(3):338–47.
[4] Faltings B, Hua KF, Schmitt G, Shih SG. Case-based representation of architectural design knowledge. Proc DARPA Case-Based Reasoning Workshop 1991;307–16.
[5] Fenves SJ, Rivard H, Gomez N. SEED-Config: a tool for conceptual structural design in a collaborative building design environment. Artif Intell Engng 2000;14:233–47.
[6] Flemming U. In: Carrara G, Kalay YE, editors. Case-based design in the SEED system knowledge-based computer-aided architectural design. Amsterdam: Elsevier; 1997.

[7] Friedman M. Essays in positive economics. Chicago, IL: University of Chicago Press; 1953.
[8] Gonzalez AJ, Lingli X, Gupta UM. Validation techniques for case-based reasoning systems. IEEE Trans Syst Man Cybernet—Part A: Syst Humans 1998;28(4):465–77.
[9] Karshenas S, Tse J. A case-based reasoning approach to construction cost estimating. Congress Comp Civil Engng 2002;113–23.
[10] Kolodner J. Case-based reasoning. San Mateo, CA: Morgan Kaufmann; 1993.
[11] Kriegsman M, Barletta R. Building a case-based helpdesk application. IEEE Intell Syst 1993;8(6):18–26.
[12] Kumar H, Krishnamoorthy C. A framework for case-based reasoning in engineering design. Artif Intell Engng Des Anal Manufact 1995;9:161–82.
[13] Morcous H, Rivard AM, Hanna AM. Case-based reasoning system for model-ing infrastructure deterioration. J Comp Civil Engng 2002; 16(2):104–14.
[14] Quinlan JR. Introduction to decision trees. Mach Learning 1986;1(1): 81–106.
[15] Reich Y, Fenves SJ. System that learns to design cable-stayed bridges. J Infrastruct Syst 1995;121(7):1090–100.
[16] Rivard H, Fenves SJ, Gomez N. Case-based reasoning for conceptual building design. First Int Conf New Information Technol Decision Making Civil Engng, Montreal, Canada 1988;355–66.
[17] Shih SG. In: Schmitt GN, editor. Case-based representation and adapta-tion in design in CAAD futures' 91. Braunshweig: Viewleg; 1991. p. 301.
[18] Smith SD. Concrete placing analysis using discrete-event simulation. Proc Inst Civil Engrs, Struct Buildings 1998;128(4):351–8.
[19] Tah JHM, Carr V, Howes R. Information modeling for case-based construction planning of highway bridge projects. Adv Engng Software 1999;30:495–509.
[20] Watson I. Applying case-based reasoning: techniques for enterprise systems. San Mateo, CA: Morgan Kaufmann; 1997.
[21] Yau N-J, Yang J-B. Case-based reasoning in construction management. Comput-Aid Civil Infrastruct Engng 1998;13(2):143–50.

Lognormal Distribution Provides an Optimum Representation of the Concrete Delivery and Placement Process

L. Darren Graham[1]; Simon D. Smith[2]; and Paul Dunlop[3]

Abstract: The process of concrete supply and delivery exhibits traits of random nature, which render control arduous. This random nature allows concrete placement operations to be considered as a stochastic system and therefore cannot be analyzed by deterministic techniques. In modeling the process realistically, it may be necessary to recreate the variability through fitting a sufficiently flexible theoretical probability distribution to observed data. There is a significant body of work relating to the probability distributions which represent general construction activities. However, there is a lack of literature aimed at determining the optimum representative of the concrete placement process. Therefore, the aim of this paper is to identify a sufficiently flexible representative of this process. To this end, construction data were collected from a number of concrete pours in Scotland, U.K. To identify a suitable distribution in this context, a computer package, *PDFit*, is presented. It is founded upon the production of probability density functions of select theoretical probability distributions plotted against the histogram of the input data. The principal method of assessment is a visual comparison of the shape of the probability density function and the input data histogram. This paper presents the lognormal distribution as a sufficiently flexible theoretical probability distribution to represent the concrete placing process.

DOI: 10.1061/(ASCE)0733-9364(2005)131:2(230)

CE Database subject headings: Concrete placing; Probability distribution; Probability density function; Stochastic processes; United Kingdom.

Introduction

The process of concrete supply and delivery to a construction site commences at a batching plant, when a truck mixer is supplied with a stipulated volume of fresh concrete. On arrival at site, the truck mixer enters "service" (if no other truck mixers are present) or joins the back of a queue of waiting truck mixers. Service requires the maneuvering of the truck into position and discharging its load into the hopper of the pump. Subsequently, the concrete is relayed into the prepared formwork. The empty truck is then washed out in a preassembled area. When clean, it joins the backcycle and returns to the batching plant.

If the process were ideal, the rate at which truck mixers arrive, position, and have their load discharged would be constant. It would then be possible to estimate deterministically the time between arrivals (interarrival time) of the trucks, to avoid queuing and underutilization of the plant. However, the process of concrete supply and delivery exhibits traits of random nature, which render control arduous. This random nature allows concrete placement operations to be considered as a stochastic system, consequently prohibiting accurate analysis by deterministic methods. This stochastic system can however be represented as a queuing system, allowing analysis by techniques such as computer simulation, neural networks, petri nets, queuing theory, and regression analysis.

The concrete supply and delivery process can be considered as two separate processes: Batching and delivery; and pump and placement. To simplify the analysis of this process, the modeler could "ignore" the batching and transport of the concrete and simply model the arrival of a truck mixer on site as a random process and considering the whole operation from the time at which the truck mixer enters the site, until its time of exit. Within this period, the following factors must be considered; interarrival time (time between the arrival of trucks), position time (time to position the truck at the pump), and pump time (time to discharge the load from the truck-mixer).

To ensure the production of a realistic queuing system model, it is necessary to recreate the variability of the concrete placing system. It is desirable to be able to represent this variability through probability distributions of sufficient flexibility. There are two methods to achieve this. They are:

- Using the observed data to define an empirical distribution, and
- Fitting a theoretical distribution to the observed data

The latter is preferable as it allows the modeler to analyze the system for situations outside the norm. This is essential when undertaking an experimental analysis.

[1]Postgraduate Student, School of Engineering and Electronics, The Univ. of Edinburgh, Edinburgh, EH9 3JL, UK.

[2]Lecturer in Project Management, School of Engineering and Electronics, The Univ. of Edinburgh, Edinburgh, EH9 3JL, UK (corresponding author). E-mail: simon.smith@ed.ac.uk

[3]Postgraduate Student, School of Engineering and Electronics, The Univ. of Edinburgh, Edinburgh, EH9 3JL, UK.

Note. Discussion open until July 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on November 12, 2002; approved on February 10, 2004. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 131, No. 2, February 1, 2005. ©ASCE, ISSN 0733-9364/2005/2-230–238/$25.00.

Probability distributions have been used by numerous researchers in the modeling of construction activities. In the study of earthmoving operations; Al-Masri (1985) used the lognormal distribution, Farid and Aziz (1993) used the beta distribution, O'Shea et al. (1964) used the exponential erlang distribution, Clemmens and Willenbrock (1978) used the erlang and normal distributions, Farid and Koning (1994) used the beta distribution, Smith et al. (1996) used the erlang distribution and Carmichael (1987) used the erlang distribution. Abourizk et al. (1994) fitted the beta distribution to all construction activities, Abourizk and Halpin (1992) used the beta distribution to model input data for construction simulation, and Maio et al. (2000) carried out an extensive survey of the distributions which represent construction data. Finally, Smith (1998, 1999) experimented with the applicability of the erlang, Weibull, gamma, and beta distributions to concrete placement data.

It is essential in simulation applications that the selected distribution to model the collected data truly reflects the properties of the data (Abourizk and Halpin 1990). Therefore, it is of concern that there is a lack of literature aimed at determining if the probability distributions deemed appropriate for use in construction activities, generally, are suitable for use as representatives of the concrete placement process. Consequently, the aim of this paper is to determine a probability distribution, which is sufficiently flexible to represent the concrete placement process.

A primary conclusion of this paper is that the lognormal distribution is a sufficiently flexible theoretical distribution to represent the concrete placing process. The selected distribution can be used to create random variates for use in a simulation model of the process. This model, upon validation, can be used to yield estimating and planning information, which could improve the performance of civil engineering construction companies in actual concrete placement operations.

PDFit Program

This paper presents a new computer application, *PDFit*, which was developed using Visual Basic for Applications programming language, embedded into the Microsoft Excel spreadsheet package. It is based on the production of probability density functions (PDFs) of select theoretical probability distributions plotted against the histogram of the input data. The distributions that have been utilized in *PDFit* are the uniform, normal, lognormal, exponential, Weibull, gamma, and beta. The mathematical forms of these distributions are provided by Law and Kelton (1991). Actual construction data were collected from large concrete pours on four construction projects in Scotland, U.K. This data provided the input to the *PDFit* program. The principal method of assessment was a visual comparison of the shape of the PDF and the underlying histogram. However, formal statistical methods are employed in *PDFit* to provide mathematical espousal for the visual assessment.

The steps involved in *PDFit* to fit theoretical probability distributions to the input data are:

1. Create an input data histogram,
2. Estimate parameters,
3. Plot PDF(s) against input data histogram,
4. Visually assess the fit, and
5. Conduct goodness-of-fit tests.

Data Collection

The raw data collected for this investigation were the pour date, weather, target slump, location of pour, batch time, arrival time, start and completion time of pumping, batch quantity, concrete slump, truck wait time, and operation inactive time.

The data were collected from two major construction projects in the United Kingdom discussed below.

Project 1. Aberdeenshire Wastewater Treatment Plant

The Aberdeenshire wastewater project constitutes the design, build, operate, and maintenance of four treatment plants. These will serve 250,000 customers in Aberdeen, Stonehaven, Peterhead, and Fraserburgh in the north of Scotland, under a 30-year concession. The financing of this £80 million project has been organized under the United Kingdom government's Private Finance Initiative. The data collected were from the Aberdeen, Fraserburgh, and Peterhead sections of the project. A fourth section, Stonehaven, was not included as construction activities had not commenced during the period of data collection.

From the Aberdeen site, 154 concrete pours were sampled. The volume of the samples ranged from 12 to 294 m^3, with an average of 57.1 m^3. The average number of truckloads and the average delivery volume per pour were 7 and 7.33 m^3, respectively. From the Fraserburgh site, 20 concrete pours were sampled. The volume of the samples ranged from 32 to 315 m^3, with an average of 125.9 m^3. The average number of truckloads and the average delivery volume per pour were 19 and 7.16 m^3, respectively. From the Peterhead site, 10 concrete pours were sampled. The volume of the samples ranged from 18 to 294 m^3, with an average 143.9 m^3. The average number of truckloads and the average delivery volume per pour were 19 and 7.2 m^3, respectively.

Project 2. Tay Wastewater Project

The aim of this project is to design, procure, construct, commission, and performance test a wastewater treatment plant near Dundee, U.K. The plant will serve customers in Dundee, Carnoustie, and Arbroath. The financing of the project is being organized under the United Kingdom government's Private–Public Partnership Initiative.

The data collected were from the Arbroath site, where 48 concrete pours were sampled. The volume of the samples ranged from 12 to 180 m^3, with an average of 76.4 m^3. The average number of truckloads and the average delivery volume per pour were 12 and 5.89 m^3, respectively. In total, 232 pours were sampled and used in this investigation.

Upon completion of data collection, the raw data were entered into a Microsoft Excel spreadsheet and the values of interest were derived from the raw data (an example is shown in Table 1). These values were used as the input to the *PDFit* program, to allow the investigation of which theoretical probability distribution is a suitable representative of the concrete placing process.

Fitting Distributions

Creating an Input Data Histogram

To make a histogram, the range of values covered by the data is broken up into k disjoint adjacent intervals $[b_{0(\text{zero})}, b_1]$,

Table 1. Sample of Input and Derived Data from Real Construction Sites

	Observed time (h:min)				Quantity			Derived time (h:min)					
Truck	Batch	Arrive	Start	Complete	Batch	Cumulative	Slump	Interarrival time	Truck queue time	Truck position time	Pump time	Pump rate (m^3/min)	Pump idle time
N584	10:55	11:12	11:17	11:43	8	8	85				0:26	0.31	Truck on time
T662	11:10	11:28	11:46	12:01	6	14	75	00:16	00:15	00:03	0:15	0:40	Truck on time
H921	11:45	11:59	12:04	12:34	8	22	95	00:31	00:02	00:03	0:30	0.27	00:37
H922	12:55	13:11	13:15	13:29	6	28	95	01:12	Truck late	00:41	0:14	0.43	Truck on time
T662	13:10	13:27	13:32	13:47	6	34	85	00:16	00:02	00:03	0:15	0.40	00:17
H921	13:45	14:04	14:06	14:25	8	42	85	00:37	Truck late	00:19	0:19	0.42	00:10
H922	14:20	14:35	14:38	14:50	6	48	100	00:31	Truck late	00:13	0:12	0.50	
	Overall operation time 3 h 55 min												
				Averages	6.86		88.57	00:33	00:06	00:13	0:18		00:21

Note: Pour date: August 3, 2000; weather: O/cast & cool; target slump: 75; and location: aeration tank wall pour 5.

$(b_1,b_2),\ldots,(b_{k-1},b_k)$. All intervals should be the same width, say, $\Delta b=b_j-b_{j-1}$, which might necessitate removing extremely large or small X_i's to avoid producing an unwieldy looking histogram plot. For $j=1,2,\ldots,k$, let h_j be the proportion of the X_i's that are in the jth interval (b_{j-1},b_j). Finally, the function is defined as

$$h(x)=\begin{cases}0_{(\text{zero})} & \text{if } x<b_0\\ h_j & \text{if } b_{j-1}\le x\le b_j \text{ for } j=1,2,\ldots,k\\ 0_{(\text{zero})} & \text{if } b_k\le x\end{cases} \quad (1)$$

and it is plotted as a function of x. The plot of h, which is piecewise constant, is then compared with plots of densities of various distributions on the basis of shape alone (location and scale differences are ignored).

Histograms are applicable to any continuous distribution and provide a readily interpreted visual synopsis of the data. Furthermore, it is easy to visually judge the closeness of fit of density functions to the histogram. There are, however, certain difficulties with using histograms. Most irksome is the absence of a definitive guide for choosing the number of intervals, k (or equivalently, their width Δb).

The research literature contains numerous views on the optimum method of applying class interval rules. Montgomery and Runger (1994) proposed, in order to avoid uninformative histograms with too many or few class intervals, between 5 and 20 intervals should be used. They suggested that it is more accurate to calculate the number of intervals by taking the square root of the number of observations. Their view is supported by Kottegoda and Rosso (1998), but the researchers extended the class intervals rule by setting a lower and upper bound of class interval quantity of 5 and 25, respectively. Neville and Kennedy (1964) suggested 10 and 25 as the lower and upper bounds, respectively. Alternatively, an equitable allocation of the frequency is obtained by dividing the total area under the PDF into equal areas. This is the equal-probabilities method of constructing classes, Mann and Wald (1942). This method was later clarified by Williams (1950) and Hogg and Ledolter (1992). Law and Kelton (1991) proposed that a number of class intervals, k, should be used to produce a number of histograms. The histogram that has the smoothest shape should be utilized further. The writers decided upon consideration of the research literature, to select a range of class intervals, k, for which histograms were plotted. They were: 8, 10, 12, and 14 class intervals. For each of these intervals, a histogram was plotted and a visual comparison was undertaken to ascertain which histogram had the smoothest shape.

Parameter Estimation

Theoretical distributions have at least two parameters which define them (Smith 1998). There are a number of methods which can be utilized to estimate these parameters from the input data, such as, maximum likelihood estimators (MLEs), moment matching, and least-squares estimates. However, this discussion shall consider only MLEs further, as this method has been generally acknowledged to provide the most satisfactory estimates (Law and Kelton 1991).

The likelihood function is

$$L(\theta)=f_\theta(X_1)\bullet f_\theta(X_2)\cdots f_\theta(X_n) \quad (2)$$

The MLE is defined to be the value of θ that maximizes $L(\theta)$ over all permissible values of θ.

Although MLEs are widely regarded as the method which provides the most satisfactory parameter estimates, there has been criticism in recent research literature. Abourizk et al. (1994) criticized the use of the MLE method because it requires prior knowledge of the endpoints of the beta distribution, and that this information is not commonly known in construction processes. Further criticism was that this method is extremely sensitive to the values of L and U. Abourizk et al. (1994) recommended that the method should only be used by a simulation analyst with extensive knowledge of the properties of the underlying distribution to verify the endpoints.

Plotting Probability Density Functions Against the Input Data Histogram

The estimated parameters of a distribution allow the PDF to be calculated. This PDF is plotted as a function of x.

There are many theoretical distribution functions which could fit a set of sample data, and because of this, the goodness of fit between the theoretical function and data should be assessed (Maio et al. 2000). One such method is a visual assessment. To allow a visual assessment of the closeness of fit to be undertaken, the PDF plot must be placed on top of the plot of the input data histogram. Fig. 1 shows a plot of the select PDFs against the interarrival time data histogram.

Assessment of Fit

To undertake a *visual* assessment of the goodness of fit, the shape of each PDF should be compared with the shape of the histogram.

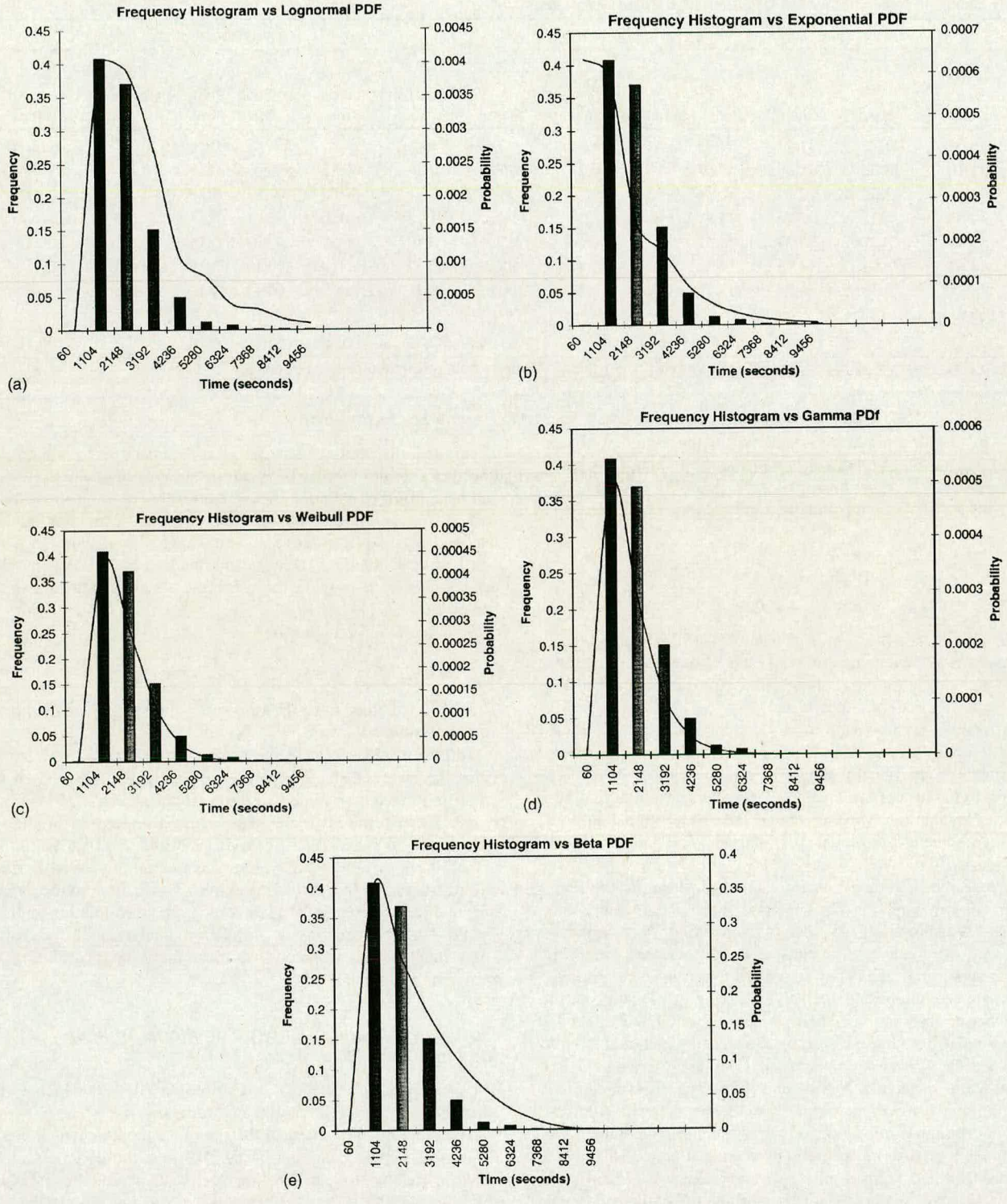


Fig. 1. Lognormal, exponential, Weibull, gamma, and beta, probability density functions for interarrival time

The *PDFit* program places the plots of the PDF against the histogram for each distribution together, to aid the comparison process. The PDF that has the closest matching shape, should be selected and used to represent the input data in any further analysis.

Abourizk et al. (1991) expressed the opinion that visual assessment of the fit is often the superior method of assessing the closeness of fit. However, visual assessment will always be subjective and prone to potential human error and therefore mathematical support for the assessment can improve confidence in the selection. *PDFit* includes two formal statistical tests to provide such mathematical support, they are: Kolmogorov–Smirnov

(K–S) and Anderson–Darling (A–D) tests. An alternative formal statistical assessment of fit is the Chi-square test. Although the Chi-square test was used by Clemmens and Willenbrock (1978), it has not been included in *PDFit* because the K–S and A–D tests have several advantages:

- There is no requirement for grouping of data,
- The test results are valid for the exact sample size, and
- The tests are more powerful in comparing alternative distributions.

Kolmogorov–Smirnov (K–S) Test

The K–S test compares the observed data with the hypothesized theoretical distribution. This test does not require the grouping of data and eliminates the problem of interval specification (this being a problem with the Chi-square test, Maio et al. 2000). It also checks if the observed data could have originated from a theoretical distribution, with the estimated parameters (Maio et al. 2000).

$$F_n(x) = \frac{\text{number of } X_i's \leq x}{N} \text{ for all real numbers } x \qquad (3)$$

Thus, $F_n(x)$ is a (right-continuous) step function such that

$$\times F_n(X_{(i)}) = \frac{i}{n} \qquad (4)$$

for $i = 1, 2, \ldots, n$. If $F(x)$ is the fitted distribution function, a natural assessment of goodness of fit is a measure of the closeness between $F_n(x)$ and $F(x)$. The K–S test statistic D_n is simply the largest (vertical) distance between $F_n(x)$ and $F(x)$ for all values of x and is defined formally by

$$D_n = \sup_x\{|F_n(x) - F(x)|\} \qquad (5)$$

Abourizk et al. (1991) expressed the opinion that the K–S statistic can be used for assessing the quality of fit from various methods of statistical analysis, e.g., fitting theoretical distributions, empirical distributions, etc. Abourizk et al. (1994) used the K–S test on construction process data sets.

Anderson–Darling Test

The A–D test differs from the K–S test on the weight given to the tails of the distributions (Maio et al. 2000). In fact, the A–D test is designed to detect discrepancies in the tails of probability distributions (Law and Kelton 1991).

The A–D test statistic A_n^2 is defined by

$$A_n^2 = n\int_{-\infty}^{\infty} [F_n(x) - \hat{F}(x)]^2 \Psi(x)\hat{f}(x)dx \qquad (6)$$

Thus, A_n^2 is just the weighted average of the squared differences $[F_n(x) - F(x)]^2$ and the weights are the largest for $F(x)$ close to 1 (right tail) and $F(x)$ close to 0 (left tail).

It was suggested that the A–D test has higher power than the K–S test against many alternative distributions (Stephens 1974). The A–D test has not, to date, been used extensively in construction-based research (Maio et al. 2000).

Experimental Analysis and Results

The aim of the experimental analysis was to identify a theoretical probability distribution that is representative of the concrete placement process. To achieve this aim, two experiments were devised.

Experiment One: Combined Concrete Pours

The differences in the experiments lie in the data which were selected to be inputted into *PDFit*. Experiment 1 involved combining the data which were gathered from real concrete pours, into large data sets. The main factors of interest, the interarrival, position, and pump times constituted 2,180, 2,177, and 2,364 time durations, respectively. Although these data sets would prove unrepresentative of an individual concrete pour, the results of Experiment One will provide a general overview of the concrete placement process.

Interarrival Time Results

The comparison plot for the lognormal, exponential, Weibull, gamma, and beta distributions are shown in Figs. 1(a–e). The smoothest histogram contained 10 class intervals. The histogram of the input data is positively skewed (to the right) and has a right tail which swiftly falls in probability. Due to the nature of this histogram, the uniform and normal PDFs are ruled out immediately from further consideration, as they are of a symmetrical disposition. The remaining distributions all bear some resemblance to the underlying histogram and require a more detailed inspection. Although the Weibull [see Fig. 1(c)], gamma [see Fig. 1(d)], and beta [see Fig. 1(e)] PDFs skew in the correct manner in relation to the underlying histogram, their right tails fall with a shallow gradient compared with that of the histogram. This results in an overestimation of the input data. The exponential PDF [see Fig. 1(b)] is ruled out because it provides an underestimation of the input data, which are most prevalent in the tail of the distribution. The importance of achieving an accurate representation of the input data when simulating a process cannot be overemphasized. Therefore, any over- or underestimation will reduce the accuracy of the analysis.

Therefore, the lognormal distribution [see Fig. 1(a)] is considered to be the optimal representative of interarrival time, subject to results of goodness-of-fit tests. The results of the K–S and A–D tests (see Table 2) provide statistical support for the conclusion drawn in the visual inspection, that the lognormal distribution best represents the interarrival time data.

Position Time Results

The comparison plots for select distributions are shown in Figs. 2(a)–(d). The smoothest histogram contained 12 class intervals. The underlying histogram is significantly skewed to the right, which will result in any symmetrical distributions (i.e., the uniform and normal) being considered unrepresentative of the input data. The underlying histogram has a right tail which falls in an exponential manner. This nature allows the exclusion of the beta distribution, as it provides an inferior fit compared with the remaining distributions. The selection is now between the lognormal [see Fig. 2(a)], exponential [see Fig. 2(b)], Weibull [see Fig. 2(c)], and gamma [see Fig. 2(d)] distributions.

Upon close inspection of the remaining curves, the lognormal distribution appears to provide the optimum representation of the input data. The exponential, Weibull, and gamma distributions all attain a similar shape to the underlying histogram, but underestimate the input data, notably in the tail of the distribution. The statistical test (see Table 2) results support the visual assessment, ranking highly the exponential and Weibull PDFs. The K–S test

Table 2. Results for Experiment One: Distributions Ranked in Ascending Order

Test	Uniform	Normal	Lognormal	Exponential	Weibull	Gamma	Beta
Kolmogorov–Smirnov							
Interarrival time	7	6	2	1	5	3	4
Potential time	7	5	4	1	3	2	6
Pump time	7	3	1	2	4	6	5
Anderson–Darling							
Interarrival time	7	6	1	5	3	4	2
Position time	7	5	1	2	3	4	5
Pump time	7	5	1	6	3	2	4

results (see Table 2) contrast with those of the A–D test, regarding the ranking of the lognormal and gamma distributions.

Pump Time Results

A visual assessment was carried out for pump time data in the same manner, as that for interarrival and position times. The conclusions drawn by the writer were that the lognormal and exponential PDFs provide acceptable representation of the input data. However, it was felt that the lognormal provided better estimates of the shorter and longer pump times. The formal statistical tests (see Table 2) confirmed that the lognormal was the optimal distribution to represent pump times.

Discussion on Experiment One

There is a discrepancy between the results of the visual assessment and K–S test, for interarrival and position times. For position time, the exponential distribution is favored by the K–S test, but can clearly be seen by visual means, to provide an inferior fit in the tail of the distribution. Therefore, the A–D test result is probably more significant in this case, as the test is designed to detect discrepancies in the tails of distributions (Law and Kelton 1991). Furthermore, the A–D test is reported to have higher power that the K–S test against many alternative distributions (Stephens 1974).

Through consideration of the results of Experiment One, for each of the duration types, it can be seen that the lognormal distribution is best at representing the concrete placement process. However, the data sets used in this experiment were large and would prove unrepresentative of individual concrete pours, therefore, a second experiment has been undertaken to investigate if the lognormal distribution is representative of individual pours.

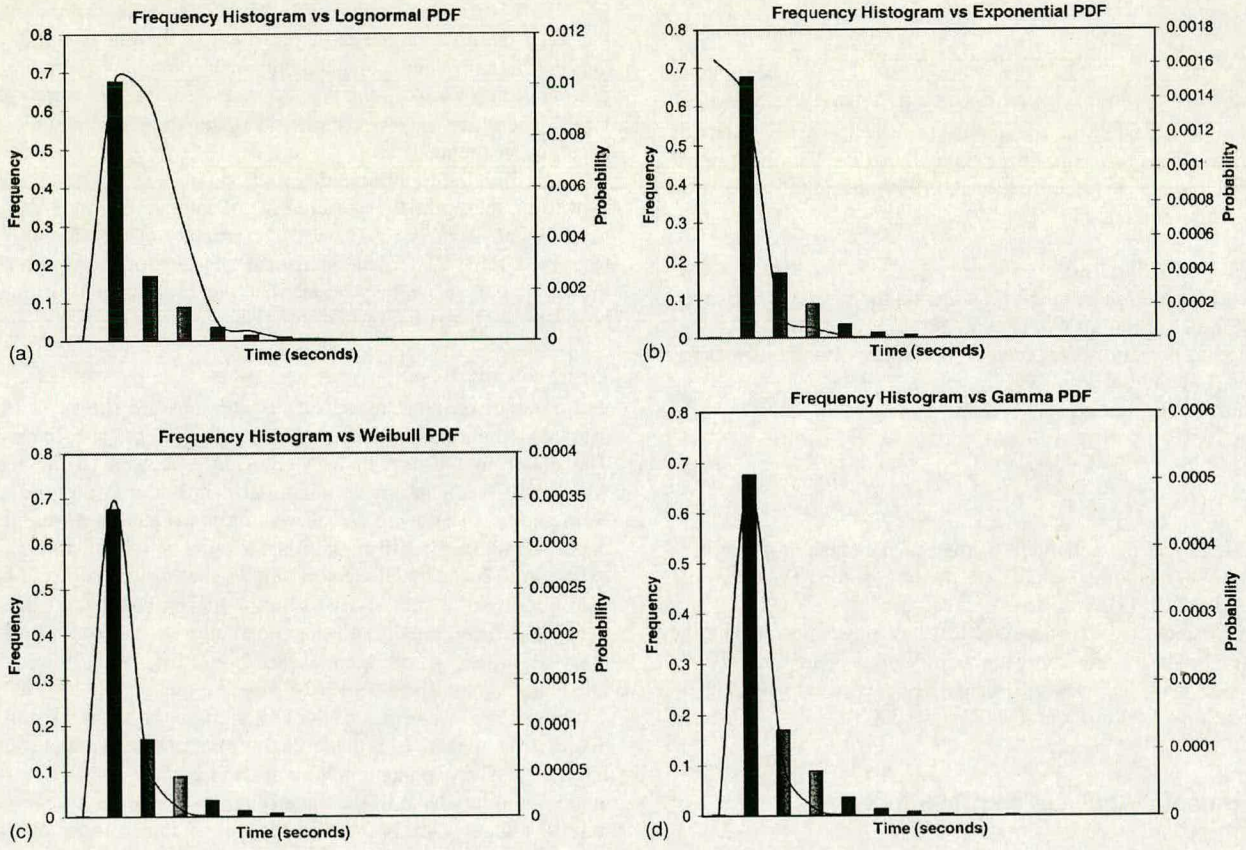


Fig. 2. Lognormal, exponential, Weibull, and gamma probability density functions for position time

Table 3. Percentage of First Ranked Distributions

Time	Uniform	Normal	Lognormal	Exponential	Weibull	Gamma	Beta
Interarrival							
Visual assessment	2.7	17.1	42.3	6.3	2.7	23.4	5.5
K–S test	0.0	8.0	25.2	0.3	0.0	20.1	46.4
A–D test	0.0	0.7	60.5	8.0	1.0	19.8	10.0
Position							
Visual assessment	2.7	6.3	42.3	7.2	2.8	17.1	21.6
K–S test	1.0	2.0	25.0	4.1	0.2	18.6	49.1
A–D test	0.0	0.0	55.4	5.4	0.6	18.4	20.2
Pump							
Visual assessment	2.7	8.1	39.6	9.0	0.0	17.1	23.5
K–S test	0.0	2.0	30.0	0.5	0.0	17.2	50.3
A–D test	0.0	0.9	58.9	2.8	0.0	17.6	19.8

Note: K–S=Kolmogorov–Smirnov; A–D=Anderson–Darling.

Experiment Two: Individual Concrete Pours

The objective of Experiment Two was to determine if the result of Experiment One is representative of individual concrete pours, in general. To achieve this objective the data from 232 concrete pours were inputted individually into *PDFit*. The results from the computer program were assessed in the same manner as in Experiment One: Visual assessment, K–S, and A–D tests.

Interarrival Time Results

A visual inspection of the 232 data sets revealed that the lognormal distribution is the most representative of the input data (see Table 3 and Fig. 3).

The underlying histogram, in general, had a strong positive skew (to the right) and the lognormal and gamma distributions are good representatives. This view is supported by the results of the visual assessments, as the lognormal and gamma distribution were ranked first in 42.3 and 23.4% of cases, respectively (see Table 3 and Fig. 3).

The statistical tests provide some support for the results of the visual assessment. The K–S test found the lognormal and gamma distributions to optimally represent the input data in a number of cases (see Table 3). However, it was found that the beta was the most representative distribution, but graphically this proved to be untrue. The A–D test found the lognormal distribution to be the optimum representative of the interarrival time data (see Table 3).

Position Time Results

The underlying histograms, in general, displayed either a strong positive (to the right) or negative (to the left) skew. This nature resulted in the lognormal, gamma, and beta distributions representing the input data in most cases (see Table 3 and Fig. 4).

Through visual assessment the lognormal distribution was found to optimally represent the position time data, having been ranked first in 42.3% of cases (see Table 3 and Fig. 4). This result was supported by the A–D test result, which ranked the lognormal distribution first in 55.4% of cases (see Table 3). The beta distri-

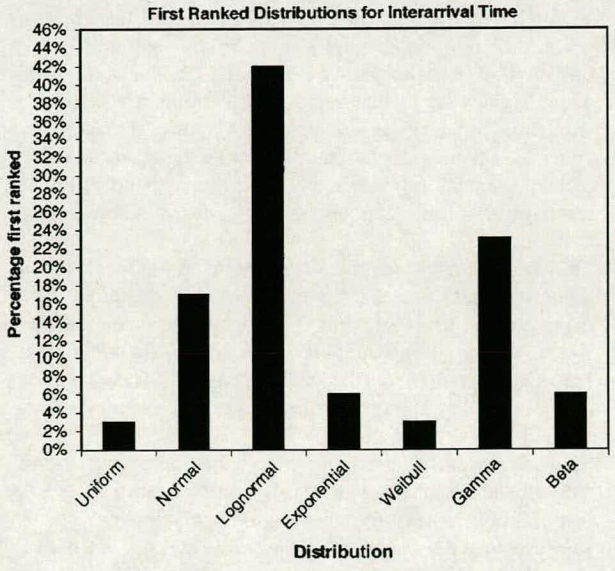


Fig. 3. Percentage of first ranked distributions using visual assessment for interarrival time

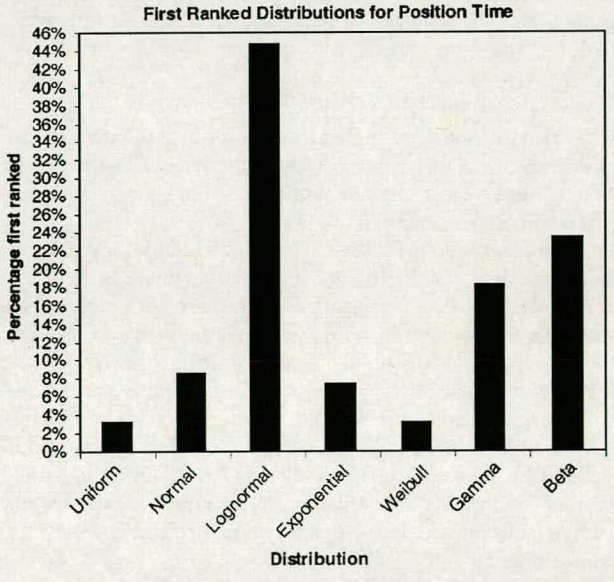


Fig. 4. Percentage of first ranked distributions using visual assessment for position time

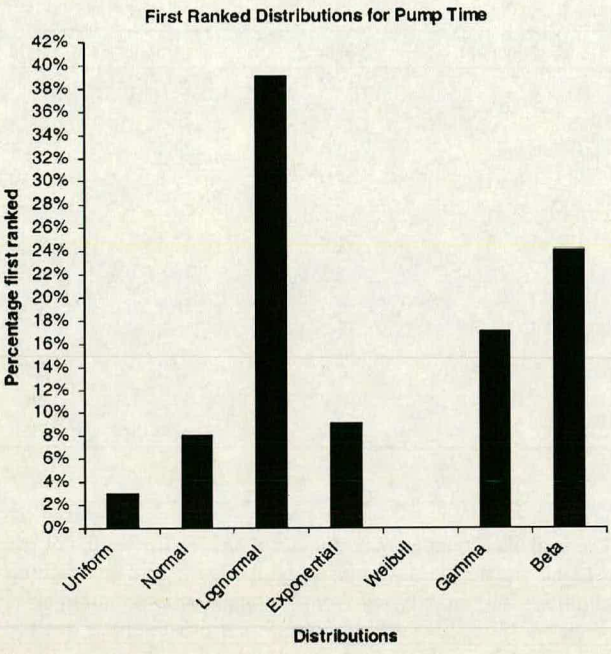


Fig. 5. Percentage of first ranked distributions using visual assessment for pump time

bution had a significant percentage of first rankings (see Table 3 and Fig. 4) because it provides the closest match to a negatively skewed histogram.

Pump Time Results

The underlying data, generally displays a range of positively and negatively skewed histograms comparable with those of the position time data. The result of the visual assessment is that the lognormal distribution is the optimum representative of the pump time data, being ranked first in 39.6% of cases (see Table 3 and Fig. 5). This conclusion is substantiated by the A–D test result (see Table 3).

Discussion on Experiment Two

Through consideration of the entire data set of 232 pours, it can be concluded that the nature of the underlying histogram was generally observed, to exhibit either a strong positive (to the right) or negative (to the left) skew.

A histogram with a positive skew was visually judged to be best represented by either the lognormal or gamma distributions. These distributions are characterized by a sharp peak close to the start of the distribution, therefore replicating the character of the skew. The lognormal distribution was visually judged to represent a larger percentage of the pours (see Table 3) than the gamma distribution. Generally, this was the due to the lognormal providing a superior fit in the tail of the distribution. This conclusion is supported by the results of the A–D test. The A–D test identifies divergence in the tails of distributions. The underlying histograms which demonstrated a negative skew were best presented by the beta distribution.

The K–S test results provided strong support for the use of the beta distribution, but this contradicts the results of the other appraisal methods. A similar conclusion to that of Experiment One can be drawn that the A–D test should take precedence over the K–S test, as it reported to have a higher power than the K–S test against many alternative distributions.

Through consideration of the experimental results the lognormal distribution has been found to be the optimum representative of the concrete placement process. This is a reflection on the evidence that underlying histograms with a positive skew were prevalent in the data set and that the lognormal distribution provides a superior fit in the tail of the distribution, to those of the other distributions. This conclusion can be reached when considering the data set combined (Experiment One) and all 232 pours, individually (Experiment Two).

Discussion and Conclusions

As previously discussed, the process of concrete supply and delivery demonstrates random qualities, which make planning, estimation, and control difficult. These traits allow concrete placement operations to be considered as a stochastic system and thus prohibits accurate analysis by deterministic methods. However, the process can be represented as a queuing system allowing analysis by nondeterministic techniques such as simulation.

The variability of the concrete placement process must be recreated through fitting a theoretical probability distribution to observed data. It is essential that the selected distribution truly reflects the properties of the input data.

Data from real construction sites in the United Kingdom were collected and provided an input to a computer program, *PDFit*. The program was developed to apply the uniform, normal, lognormal, exponential, Weibull, gamma, and beta probability distributions to the input data. The parameters of these distributions were calculated using the method of MLEs. MLEs, in general, provide the most satisfactory estimates of parameters. The User selects the best fitting distribution using a combination of visual assessment and goodness-of-fit test results.

The aim of the experimental analysis was to identify a theoretical probability distribution that is representative of the concrete placement process. The results can be summarized as follows:

1. Through visual assessment, it was determined that the lognormal distribution generally provides the optimum representation of the process. When individual concrete pours were judged visually, the lognormal distribution was the preeminent representative in a significant number of cases. The conclusion that the lognormal distribution represents the concrete placement process acquires further support from the results of the visual appraisal of large sets of concrete pours data.
2. The Weibull and gamma distributions provide close, although underestimated, representations of the large input data sets used in Experiment One. The importance of achieving an accurate representation of the input data when simulating a process cannot be overemphasized. Therefore, any over- or underestimation will reduce the accuracy of the analysis.
3. The goodness-of-fit tests provided some support to the results of the visual assessments. In some instances, the K–S test, highly ranked the distributions which were judged (through visual assessment) not to closely represent the observed data. It can be concluded that while the K–S test does not provide wholly inaccurate results, the fact that it does not place any emphasis on the tail of a probability distribution is

likely to be significant with the construction distributions under investigation. The underlying histogram for concrete placing data has, in general, a tail (right or left) which is a prominent feature. Therefore, a test which emphasizes tail weighting was more appropriate to assess the goodness of fit. Such a test is the A–D. For individual pours the A–D test ranked highly the lognormal distribution, and this corresponds with the results of the visual assessment.

This paper has presented the lognormal distribution as a sufficiently flexible theoretical probability distribution to represent the concrete placing process. The selected distribution can be used to create random variates for use in a simulation model of the process. The generation of lognormal random variates can be attained efficiently through the use of an acceptance–rejection method as described in Law and Kelton (1991). The model, upon validation, can be used to yield estimating and planning information, which could improve the performance of civil engineering construction companies in actual concrete placement operations.

Notation

The following symbols are used in this paper:

$An2$ = Anderson–Darling test statistic;
Dn = Kolmogorov–Smirnov test statistic;
$F(x)$ = fitted distribution function;
$F_n(x)$ = empirical distribution function from the data set, $X_1, X_2, \ldots, X_n$;
$F_\theta(x)$ = hypothesized density functions, assuming one unknown parameter, θ;
L = lower limit of distribution;
U = upper limit of distribution;
sup = set of random numbers A is the smallest value that is greater than or equal to all members of A; and
$\Psi(x)$ = weight function $= \Psi(x) = 1/\{F(x)[1-F(x)]\}$.

References

Abourizk, S. M., and Halpin, D. W. (1990). "Probabilistic simulation studies for repetitive construction processes." *J. Constr. Eng. Manage.*, 116(4), 575–594.

Abourizk, S. M., and Halpin, D. W. (1992). "Modeling input data for construction simulation." *Proc., 8th National Conf. on Computing in Civil Engineering*, ASCE, Reston, Va., 1147–1154.

Abourizk, S. M., Halpin, D. W., and Wilson, J. R. (1991). "Visual interactive fitting of beta distributions." *J. Constr. Eng. Manage.*, 117(4), 589–605.

Abourizk, S. M., Halpin, D. W., and Wilson, J. R. (1994). "Fitting beta distributions based on sample data." *J. Constr. Eng. Manage.*, 120(2), 288–305.

Al-Masri, F. M. (1985). "Analysis of a loader–truck system operation using the computer simulation language *SIMAN*." Pennsylvania State Univ., University Park, Pa.

Carmichael, D. G. (1987). "A refined queuing model from earthmoving operations." *Civ. Eng. Syst.*, 4, 153–159.

Clemmens, J. P., and Willenbrock, J. H. (1978). "The *SCRAPESIM* computer simulation." *J. Constr. Div., Am. Soc. Civ. Eng.*, 104(4), 419–435.

Farid, F., and Aziz, T. (1993). "Simulating paving fleets with nonstationary travel." *Proc., Computing in Civil and Building Engineering*, Vol. 2, ASCE, New York, 1198–1206.

Farid, F., and Koning, T. L. (1994). "Simulation verifies queuing program for selecting loader–truck fleets." *J. Constr. Eng. Manage.*, 120(2), 386–404.

Hogg, R. V., and Ledolter, J. (1992). *Applied statistics for engineers and physical scientists*, Macmillan, New York.

Kottegoda, N. T., and Rosso, R. (1998). *Statistics, probability, and reliability for civil and environmental engineers*, McGraw-Hill, Singapore.

Law, A. M., and Kelton, W. D. (1991). *Simulation modeling and analysis*, 2nd Ed., McGraw-Hill, New York.

Maio, C., Schexnayder, C., Knutson, K., and Weber, S. (2000). "Probability distribution functions for construction simulation." *J. Constr. Eng. Manage.*, 126(4), 285–292.

Mann, H. B., and Wald, A. (1942). "The choice of the number and width of a number of class intervals in the applications of the Chi-square tests of goodness-of-fit." *J. Am. Stat. Assoc.*, 45, 47–86.

Montgomery, D. C., and Runger, G. C. (1994). *Applied statistics and probability for engineers*, Wiley, New York.

Neville, A. M., and Kennedy, J. B. (1964). *Basic statistical methods for engineers and scientists*, International Textbook Co., Scranton, Pa.

O'Shea, J. F., Slutkin, G. N., and Shaffer, L. R. (1964). "An application of the theory of queues to the forecasting of shovel-truck fleet production." *Constr. Res. Ser. No. 3*, Dept. of Civil Engineering, Univ. of Illinois, Urbana, Ill.

Smith, S. D. (1998). "Concrete placing analysis using discrete-event simulation." *Proc. Inst. Civ. Eng., Struct. Build.*, 128(4), 351–358.

Smith, S. D. (1999). "Modeling and experimentation of the concrete supply and delivery process." *Civ. Eng. Environ. Syst.*, 16(2), 93–114.

Smith, S. D., Osbourne, J. R., and Forde, M. C. (1996). "The use of a discrete-event simulation model with Erlang probability distributions in the estimation of earthmoving production." *Civ. Eng. Syst.*, 13(1), 25–44.

Stephens, M. A. (1974). "EDF statistics for goodness of fit and some comparisons." *J. Am. Stat. Assoc.*, 69, 730–737.

Williams, C. A., Jr. (1950). "The choice of a number of class intervals in the application of the Chi-square test." *Ann. Math. Stat.*, 13, 306–317.

Modeling the Ready Mixed Concrete Delivery System with Neural Networks

L. Darren Graham[1], Doug R. Forbes[2] and Simon D. Smith[3]

[1] Postgraduate Student, School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, United Kingdom.

[2] Site Engineer, Gleeson Construction Services, Haredon House, North Cheam, Sutton, Surrey, SM3 9BS, United Kingdom

[3] Lecturer in Construction and Project Management, School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, United Kingdom. Tel: +44 131 650 7159; Fax: +44 131 650 6781. simon.smith@ed.ac.uk [*Corresponding Author*]

For *Automation in Construction*

Submission Date: 31st December 2004

Abstract

The ready mixed concrete delivery system is a common construction process in a very wide range of construction projects. The ability of the planners and estimators of such projects to accurately determine the level of resources needed, and to estimate the output of an efficient and effective operation is highly important and thus modeling of the process can be useful. This paper presents a Neural Network methodology to the modeling problem and outlines the two main architectures employed: a feed-forward network and an Elman network. Many combinations of layers, training algorithms, number of neurons, activation functions and format of data were considered and the results were validated using an independent validation data set with five goodness-of-fit tests. The results indicate that two- and three-layer feed-forward networks provide the best estimates of concrete placing productivity and that the Elman network, not previously considered in this type of study, was less successful.

Keywords: Ready mixed concrete; Neural networks; Delivery; Productivity

Introduction

This study is concerned with modeling the ready mixed concrete delivery system (RMCDS) where the concrete is pumped into its final position. Ready mixed concrete (RMC) is an essential material in contemporary construction and engineering projects and thus it is imperative that the process of acquiring and handling RMC is managed with the utmost efficiency and accuracy. Further, RMC must be delivered in a workable state and, given that it has a limited shelf-life of roughly one and a half hours, (American Society for Testing and Materials, 2003) any mismanagement of the processes of delivery and use of this vulnerable material could result in this shelf-life being exceeded. Two possible consequences of this are: the out-of-date material may be used by a construction manager under pressure to deliver a project on time, resulting in a poor quality structure; or, the contractor (or client, depending on the contract type) pays for the material (if they are at fault) and for the gap in the construction schedule, with resources standing idle awaiting the next concrete delivery. Neither of these consequences is desirable for a construction contractor working on a tight budget, and given that the above problems can be caused by inaccurate estimates of the resources required being made by practitioners, there is a need for a model of the RMCDS which has the capability of providing such estimates to a sufficient accuracy. Given a set of specific operation details e.g. number of concrete pumps, concrete pour volume, etc., such a model aims to accurately predict:

- the total operation duration
- the rate of delivery (productivity) of the RMC
- the cost of an operation
- the utilization rates of the plant (concrete pump and truck-mixers)

These estimates can be used to optimally plan an operation and its resources and to help regulate the use of RMC on-site, thereby avoiding the 'shelf-life' issue discussed above.

Any model that accurately portrays the likely cycle time of a concrete truck-mixer will also prove useful for the management of a concrete batching plant - this does not appear to have been discussed in previous publications.

How can the RMCDS be modeled? Currently in practice, an experienced planner estimates the likely rate of delivery (productivity) of RMC. So why change this practice? The data collected for this study came from real construction operations in the United Kingdom, which used experience-based estimation. An examination of this data revealed that on average 12% of the time the site-based resources (e.g. concrete pump and workers) remained idle, at an average cost to the contractor of 14% of the total cost of RMC operations. Of interest to the RMC supplier, it was found in the same data that concrete truck-mixers remained idle on-site for an average of 25% of the total duration of the concrete operations. Clearly, there is room for improvement in the RMCDS currently being used in practice.

Feed-forward neural network (NN) models have been successfully utilized in process productivity estimation problems in other areas of construction management such as earthmoving (e.g. Shi, 1999), but have not been used in concreting. It is the aim of this study to determine whether or not feed-forward NN models are capable of accurately modeling the RMCDS, to allow productivity and cycle time estimates of the process to be established. In addition, a different type of NN model, the Elman recurrent network shall be examined in the context of this study. The Elman NN has not been used in construction modeling and has rarely been used in discrete problems (of which the RMCDS is one) in other research areas. A focus has been placed upon the development of these models, leaving the implementation and use of the models in a practical environment for future research. All of the models used in this study have been developed in the Matlab computing environment on a computer with: an 850MHz processor; 128 Mb RAM memory; on the Windows ME platform.

The structure of this paper is as follows:

- Review of relevant Neural Network literature
- The NNs to be considered in this study
- Data used in the study
- Development of the NN models
- Results from training and testing the NN models
- Validation of the NN models
- Conclusions and future work in this research area

Review of relevant Neural Network literature

There exists a significant body of work relating to the development of NN models of the productivities achievable by the labor force in construction activities. For instance, Chao and Skibniewski (1994), Kartam (1996), and Flood and Christophilos (1996) focused on earthmoving; Portas and Abourizk (1997) considered concrete formwork construction. These studies found NNs capable of modeling labor productivity of specific construction activities to a high degree of accuracy. More generally, Rowings and Sonmez (1996), Abourizk and Wales (1997) and Lu et al. (2000) found NNs capable of modeling labor productivity.

Shi (1999) adopted a broader approach to those mentioned above by modeling the whole earthmoving process, not just by the labor element of that process. This model was a feed-forward NN, and was found to provide accurate estimates of the productivity of the earthmoving process. This study aims to further the body of knowledge by modeling the RMCDS using the feed-forward NN method used by Shi, and by exploring the potential held by the Elman NN method.

An Overview of the Neural Networks (NNs) Used in this Study

This overview shall examine the NN architectures used in this study and provide a summary of the functions used to educate (train) the NN models. The

reader should refer to Haykin (1999) for a comprehensive explanation of the internal workings of NN models.

The NN architectures used in this study

There are two NN architectures being employed in this study: feed-forward and Elman recurrent. Feed-forward networks are especially useful in function approximation when a set of inputs and outputs is all that is known of the system, which is the situation in this study (Flood, 1991).

Feed-forward networks have their neurons arranged in layers. These layers have connections to the layers either side, as shown in Figure 1a. This figure shows a network with p inputs and s outputs. There are two layers of hidden neurons, 1 and 2, with q and r neurons in each respectively. In designing a feed-forward neural network, it is necessary to determine heuristically the combinations of the number of hidden layers and the number of neurons in each to obtain the optimum combination. Common notation for the number of layers in a network is described as being the number of hidden layers *plus* the output layer, since these are the layers that process the information (Hecht-Nielsen, 1990).

Figure 1a

The Elman NN, as shown in Figure 1b, has a loop from the output of the hidden layer to the input. This feedback loop allows the network to form temporal patterns (The MathWorks, 2003). By providing these temporal patterns, the network is able to see its own previous output, and so shape the responses accordingly. Elman NNs, having a recurrent loop possess memory characteristics, which allow the network not only to make use of data at a particular time step, but also the data that has been used for previous cycles (Elman, 1990). With memory capabilities, the Elman NN should provide a better generalization to the function being mapped than the feed-forward network. Thus, it is worthy of consideration in this study.

Figure 1b

The training functions used in this study

In order to produce an effective NN model, it is vital that the network is trained properly. In the training of a NN model, a function is mapped from known inputs to known outputs. This is supervised learning. In training a supervised NN, weights between the neurons are adjusted to minimize the error in the output. For this study, the Matlab Neural Network Toolbox was used to construct and train the networks. The toolbox has pre-programmed training functions. For training NNs within Matlab, the default training algorithm is the Levenberg-Marquardt (LM). However, this method requires a large amount of computing resources (Mayoraz and Vulliet, 2002). The MathWorks (2003) recommends the use of a conjugate training algorithm to increase speed. A conjugate training algorithm operates similarly to the gradient descent, but searches in conjugate directions along the gradient to produce faster convergence (Charalambous, 1992). This study considers two training algorithms: the LM because of its widespread usage; and the scaled conjugate gradient (SCG)

algorithm because it decreases NN training time (The MathWorks 2003).

Project Data: Details, Analysis and Preparation

This section aims to introduce the sources and quantity of the data used in this study. Next, the variables that were available from the collected data and those chosen for modeling shall be discussed. Finally, the model data is scaled in preparation for the training of neural network models.

Data sources and quantity

The data used in this study was taken from observation of the RMCDS on four construction projects, with unique clients and contractors, in the United Kingdom. The structures being constructed in these projects varied from wastewater treatment tanks to a major highway viaduct. The variance in client, contractor and project type has produced a diverse set of project data, and should allow this study to draw generic conclusions about the capability of neural networks to model the RMCDS.

A fundamental maxim in the development of neural networks, with respect to data set size, is the more data provided the more that a neural network can learn about that data. In this study, taking the observed data as a whole, there are 251 examples of the RMCDS. This compares favorably with previous studies modeling construction using neural networks: Rowings and Sonmez (1996) used a sample size of 112; Abourizk et al. (2001) used 39 examples; and finally, Arditi and Tokdemir (1999) used a sample size of only 12.

Table 1

The collected variables and the selection of variables to be included in modeling

There were eight variables in the collected data that were deemed suitable for use in this study (see table 1). A description of these variables may be found in Graham et al. (2004). In order to reduce this list to incorporate only the most pertinent variables a stepwise linear regression analysis was undertaken. This involved using the data in its raw form and selecting the significant variables (factors) using the t-statistic of the data. The first linear regression that was performed on the data mapped the collected variables (inputs) shown in table 1 against the productivity (output). The final productivity regression model produced the following significant variables: truck volume, total operation volume, average inter-arrival time and number of loads in an operation. Thus the original set of eight variables has been reduced to four. Four variables may appear to be too few to adequately represent the RMCDS, but if a model can be produced that requires only four variables as an input to provide an accurate estimate of productivity, then that model would be practical. Also, the linear regression model showed statistically sufficient results, with a correlation coefficient (R^2) of 0.84 – indicating that a substantial amount of the variance in the data can be described using only four variables. Further, a plot of the residuals (errors) produced by the linear regression analysis displayed in figure 2, shows that they all errors sit within 2 standard deviations and there are no clear trends in the errors. This indicates that the errors are not

correlated with the productivity. The evidence found in this study suggests that four variables are sufficient to enable models of the RMCDS to be developed.

Figure 2

Data transformation

The final step in preparing the data for use in modeling was to transform (scale) the data into the range between -1 and +1, to improve the density of the data over the problem domain. This transformation was undertaken using the formula proposed by Hegazy and Ayed (1998) which is shown in equation 1. To allow a comparison between using scaled and raw data in the training of neural networks, a set of raw data will also be used in neural network model development. The scaled and raw data sets, each consisting of four variables and 251 examples of past concreting operations are now ready for use in neural network model development.

$$Scaled\ Value = \left(\frac{2 \times (Unscaled\ Value - Varaible\ Minimum)}{Variable\ Maximum - Variable\ Minimum} \right) - 1$$

Equation 1

Development of the Neural Network (NN) Models

To re-iterate, the development of neural networks is a heuristic process. In order to investigate fully the optimum network set-up, the following characteristics were varied in the training of both network architectures:

- Number of layers
- Training Algorithm
- Number of neurons in each layer
- Activation function in each layer
- Type of data

The variations in the above network characteristics used in this study were:

- **Layers**: Networks were created with two and three layers. This corresponds to one and two hidden layers respectively.
- **Training Algorithm (TA):** The Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) training algorithms were used in this study.
- **Number of Neurons:** The number of neurons in each layer was varied from 2 to 22, in increments of 2. The limit of 22 was enforced because an unmanageable amount of interconnections in the network would result from increasing this number, and also the training time would also increase substantially.
- **Activation Function:** Three activation functions which are commonly used in neural network modeling have been used in this study. They are the tan-sigmoid (T), linear (P) and the log-sigmoid (L) activation functions. Details of the mathematical form of these functions are provided by The Mathworks (2003). NN models require that a linear (P) activation function be present in the link between the last hidden layer and the output layer. All possible combinations of

these activation functions were considered for the two and three layer NN models, resulting in 9 combinations for the two-layer NN and 27 combinations for the three-layer NN models.

- **Type of Data:** Both scaled and raw data sets were considered

Each network was trained using 75% of the 251 past examples. The remainder of the data set was used as a test set, with each network attempting to recreate (simulate) each test result. The results of this simulation were used to calculate the optimum network, based upon the mean square error (MSE) – the most common measurement of error (Leung and Chan, 2003; Shi, 2000).

Each of the network set-ups were trained and tested (developed) 4 times (except for the 3-Layer FF LM networks, which took in excess of 19 hours, and were developed only twice). From these results, the average test MSE was calculated, along with the standard deviations from this average. The standard deviations measure the ability of the networks to produce consistent results, and are therefore important indicators.

Results from Training and Testing the NN Models

The optimum results (in terms of MSE) for each number of layers, training algorithm and data type, have been recorded in table 2 for the NN models. Thousands of results have been considered to reach this stage of the investigation.

Table 2

The following conclusions may be drawn from this information:

- The effect of scaling the data is to significantly increase the NN model's MSE, but to reduce the standard deviation in that MSE. Thus, the results for scaled data are worse than for the raw data, but at least the results are consistently of poor quality. The SCG training algorithm appears to control the consistency of the output from the NN model, reducing the variation from the levels witnessed using the LM algorithm. Although consistency in the results is a significant factor, it is not the primary one – accuracy is, and thus, the scaled data models shall not be considered for further validation. In formulating an opinion on the optimum NN model accuracy and consistency shall be considered.
- There appears to be no relationship between MSE value and the activation functions used in a NN model.
- The NN models that have test MSE's within 30% of the smallest (provided by 3-layer Elman network trained using SCG) are:
    - 2-layer feed-forward, trained with LM using raw data;
    - 2-layer feed-forward, trained with SCG using raw data;

- 3-layer feed-forward, trained with LM using raw data;
- 3-layer feed-forward, trained with LM using raw data;
- 2-layer Elman, trained with SCG using raw data;
- 3-layer Elman, trained with SCG using raw data.

The capabilities of these NN model set-ups, to recreate the RMCDS shall be examined further in a validation process, using data which has neither been used in training nor testing.

Validation of the NN Models

A model can only be investigated to see how a real system would respond if the model is a valid one (Law and Kelton, 2000). Thus, the 6 NN models (see the previous section) to be considered further in this study cannot be deemed useful for practitioners, if they have not been rigorously validated.

The process of validation was undertaken using a correlated inspection approach as proposed by Law and Kelton (2000). This involved estimating the productivity of 39 real concreting operations, not used in model development, and recording the difference between the achieved and estimated productivity.

An example plot of the results from the correlated inspection is shown in figure 3. From a visual inspection of the results the 2-layer feed-forward (LM trained), 2-layer feed-forward (SCG trained) and the 3-layer feed-forward (LM trained) appear to provide the most accurate and reliable estimates of the validation concreting operations. Unfortunately, a visual comparison of all of the results was difficult to make and the results of such an appraisal cannot be declared completely objective. The results were therefore analyzed using a number of goodness of fit tests in a bid to avoid the need to make a difficult visual assessment of correlation inspection outcome, and to learn more about each model's ability to recreate the RMCDS. The results of the goodness of fit tests were ranked to allow a comparison between the NN models, based solely on predictive capability to be made (see table 3). The results from this validation process were subsequently used to determine the optimum NN model set-up.

Figure 3

Statistical Analysis of the Correlated Inspection Results

A number of goodness-of-fit measurements have been used in this study to assess which NN model set-up provided the best re-creation of the validation results. They are:

- *The D-value of the Kolmogorov-Smirnov (K-S) test.* The K-S test calculates the cumulative frequency distribution of two sets of data (validation and model simulated) and produces a plot for each set and the largest difference (D), between the two points measured. The level of significance

of D can found in Siegel and Castellan (1988) based on the size of the two samples. Generally, the smaller D is, the better the ability of the NN model to predict process productivity. The D-value of the K-S test has been used in past validation studies performed by Naylor and Finger (1967) and Siegel and Castellan (1988).

- *Theil inequality coefficient (U) developed by Theil (1958).* Please note the equation for U can be found in Theil (1958). For valid predictions, U tends towards zero. For U tending towards one, the predictions do not follow the actual pattern. U has been used in past validation studies, such as those described by Naylor and Finger (1967).
- *Root mean squared (RMS) proportional error.* This measure is a combination of: the most commonly-used error measurement in NN modeling (Twomey and Smith, 1997) – the mean squared error (MSE); and the mean prediction error (MPE), a measurement that has the advantage of measuring the proportional size of the error (Shi, 2000). The RMS proportional measurement was used by Shi (2000) in a model validation study.
- *The correlation coefficient, R^2.* A measure of the correlation between the model simulated and the validation operations can be calculated by the correlation co-efficient, R^2. The R^2 value gives an indication of the tendency of the data to lie on a 1:1 straight line when the predicted and actual values are plotted against one another, with a value tending towards 1 indicates a good fit. The correlation coefficient was used in neural network validation by Colmenares and Perez (1998) and Shi (2000).
- *The ability for 95% of the data to lie within +/- 2 standard deviations (s) of the mean.* Twomey and Smith (1997) recommended that neural network modelers should create residual plots, which are plots of the actual value against the residual; the residual being the actual output minus the predicted output. From the residual plot, the spread of the data can be seen along with any extreme values. In assessing a residual plot, generally the closer the data is situated to the mean the better the NN model's prediction capabilities are.

For each NN set-up and for each goodness-of-fit test a comparison was made between the results. There were six NN set-ups to compare and the test results were ranked from 1 (the best) to 6 (worst). These ranks were objectively assigned in the case of the D-value, U, R^2 and RMS proportional measurements, and subjectively allocated in the ability of 95% of the data to lie within +/- 2 s of the mean. For simplicity, the actual results shall not be displayed in this paper. Rather, the ranked results are shown and analyzed instead, as in table 3.

Table 3

Table 3 shows that the basic feed-forward NN models were the most able to recreate the validation data set, suggesting that they would provide the most accurate predictions of the productivities of future concreting operations. The recurrent loop in an Elman NN model makes the model have a more complex architecture than a feed-forward model. This increased complexity means the NN model is more sensitive to fluctuations in the data and missing variables from the data set. As the project data has been collected using a stopwatch operated by a human, it is likely that the data set will contain human error and that not all of the significant variables would have been included. These data collection problems may have adversely affected the prediction capability of the Elman NN models.

Conclusions and Future Work

The following conclusions may be drawn from this study:

(a) The standard NN development practice of scaling (transforming) raw data was found to be produce NN models which were of poor quality (had a high test MSE). Thus, only raw data models were considered for use in the model validation process.

(b) Feed-forward NN models were found to be able to produce productivity estimates of the RMCDS to a high degree of accuracy. This finding suggests that feed-forward NN models may be capable of modeling other similar construction processes, which as yet have not been studied.

(c) The Elman NN had not been considered before in construction process modeling, and was found to be capable of producing estimates of the RMCDS with a promising degree of accuracy

(d) Feed-forward NN models were found to be more capable of estimating the RMCDS process productivity than the Elman models. This finding is somewhat surprising given that the 3-layer Elman NN model produced the smallest MSE during training and testing, a result which should have translated to the ability to accurately estimate the process productivity, but this did not happen as successfully as it could have. The reason for this could be that the Elman architecture is more complex than a feed-forward one, as there are more interconnecting neurons, etc., and this complexity possibly makes an Elman NN more sensitive to outliers in the data and missing process variables.

Future research could be undertaken to determine some practical guidelines for use in NN modeling of construction processes. This could include information relating to: the effect of the size of the training set on model performance; the validation

methods available, and their effect on a model's acceptability; and the use of factorial analysis in analyzing a NN model.

References

AbouRizk, S. M., and Wales, R. (1997). "Combined discrete-event/continuous simulation for project planning." *J. Constr. Eng. Manage.*,123(1), 11-20.

Abourizk, S. M., Knowles, P., and Herman, U. R. (2001). "Estimating labor production rates for industrial construction activities." *J. Constr. Eng. Manage.*, 127(6), 502-511.

American Society for Testing and Materials. (2003). "*ASTM C94/C94M-04a standard specification for ready-mixed concrete.*" ASTM International, West Conshohocken, P.A., U.S.A.

Arditi, D., and Tokdemir, O. B. (1999). "Comparison for case-based reasoning and artificial neural networks." *J. of Comp. in Civ. Eng.*, 13(3), 162-169.

Chao, L-C., and Skibniewski, M. J. (1994). "Estimating construction productivity: neural-network based approach." *J. of Comp. in Civ. Eng.*, 8(2), 234-251.

Charalambous, C. (1992). "A conjugate gradient algorithm for the efficient training of artificial neural networks." *IEE Proceedings Part G*, 139, 301-310.

Colmenares, G. L., and Pérez, R. (1998). "A data reduction method to train, test and validate neural networks." *In: Proceedings IEEE Southeastcon. '98*, 24-26, April 1998, 277-280.

Elman, J. L. (1990). "Finding structure in time." *Cognitive Science*, 14, 179-211.

Flood, I. (1991). "A gaussian-based feedforward network architecture and complementary training algorithm." *In: International Joint Conference on Neural Networks*, 1991, 171-176.

Flood, I., and Christophilos, P. (1996). "Modeling construction processes using artificial neural networks." *Automation in Construction*, 4, 307-320.

Graham, L. D., Smith, S. D., and Crapper, M. (2004). "Improving construction simulation with a case-based reasoning input." *Civ. Eng. and Env. Systems*, 21(2), 137-150.

Haykin, S. S. (1999). "*Neural networks: A comprehensive foundation.*" 2nd ed., Prentice Hall Inc., New Jersey, U.S.A.

Hecht-Nielsen, R. (1990). "*Neurocomputing.*"Addison-Wesley, Wokingham, U.K.

Hegazy, T., and Ayed, A. (1998). "Neural network model for parametric cost estimation of highway projects." *J. Constr. Eng. Manage.*, 124(3), 210-218.

Kartam, N. (1996). "Neural network-spreadsheet integration for earthmoving operations." *Microcomputers in Civ. Eng.*, 11, 283-288.

Law, A. M., and Kelton, W. D. (2000). "*Simulation modeling and analysis.*" 3rd ed., McGraw-Hill, London, U.K.

Leung, C-S., and Chan, L-W. (2003). "Dual extended kalam filtering in recurrent neural networks." *Neural Networks*, 16, 223-239.

Lu, M., AbouRizk, S. M., and Hermann, U. R. (2000). "Estimating labor productivity using probability inference neural network." *J. of Comp. in Civ. Eng.*, 14(4), 241-248.

Naylor, T. H., and Finger, J. M. (1967). "Verification of computer simulation models." *Management Science*, 14(2), 92-101.

Portas, J., and AbouRizk, S. (1997). "Neural network model for estimating construction productivity." *J. Constr. Eng. Manage.*, 123(4), 399-410.

Rowings, J. E., and Sonmez, R. (1996). "Labor productivity modeling with neural networks." *In: AACE Transactions*, PRD11.

Shi, J. J. (1999). "A neural network based system for predicting earthmoving production." Constr. Manage. and Economics, 17, 463-471.

Shi, J. J. (2000). "Reducing prediction error by transforming input data for neural networks." *J. of Comp. in Civ. Eng.*, 14(2), 109-116.

Siegel, S., and Castellan, N. J. Jr. (1988). "*Nonparametric statistics for the behavioral sciences*." 2nd ed., McGraw-Hill, London, U.K.

The MathWorks. (2003). "Neural network toolbox for use with MATLAB®: user guide." http://www.mathworks.com/access/helpdesk/help/pdf_doc/nnet/nnet.pdf [last accessed 11 November (2004].

Theil, H. (1958). "*Economic forecasts and policy*." North-Holland Pub Co, Amsterdam, Holland.

Twomey, J. M., and Smith, A. E. (1997). "Validation and verification." *In: Kartam, Nabil, Flood, Ian & Garrett, James H. Jr., eds. Artificial Neural Networks for Civil Engineers: Fundamentals and Applications*. ASCE, NewYork, U.S.A., 44-64.

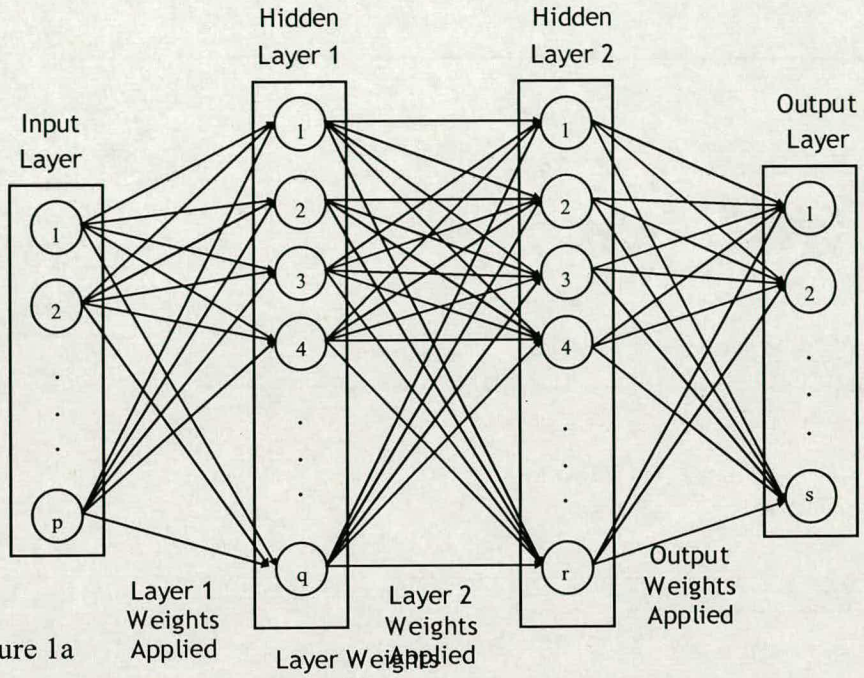


Figure 1a

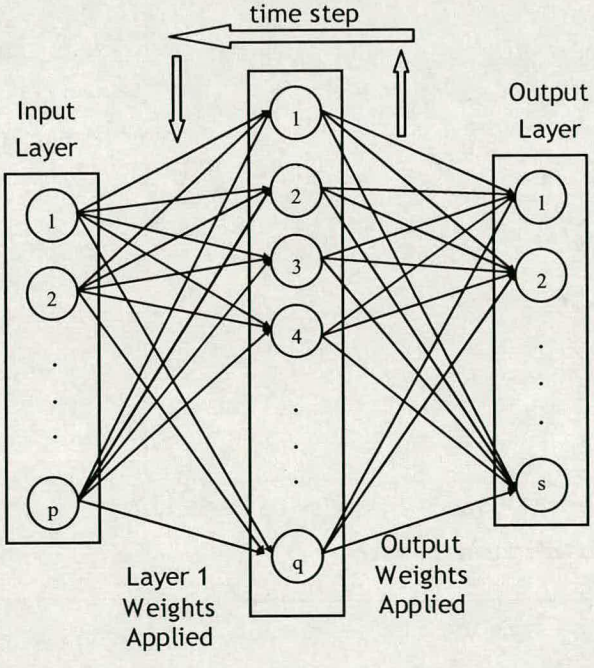


Figure 1b

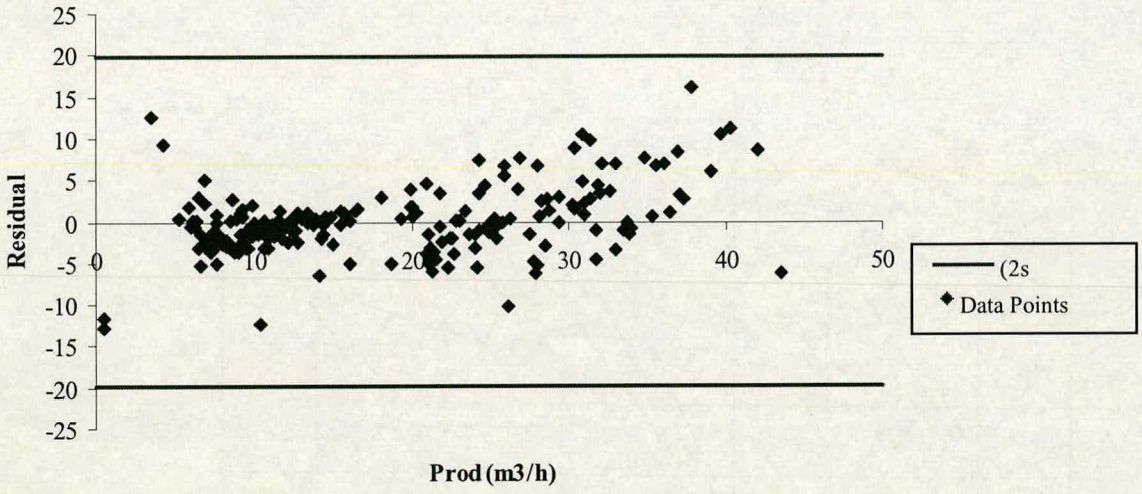


Figure 2

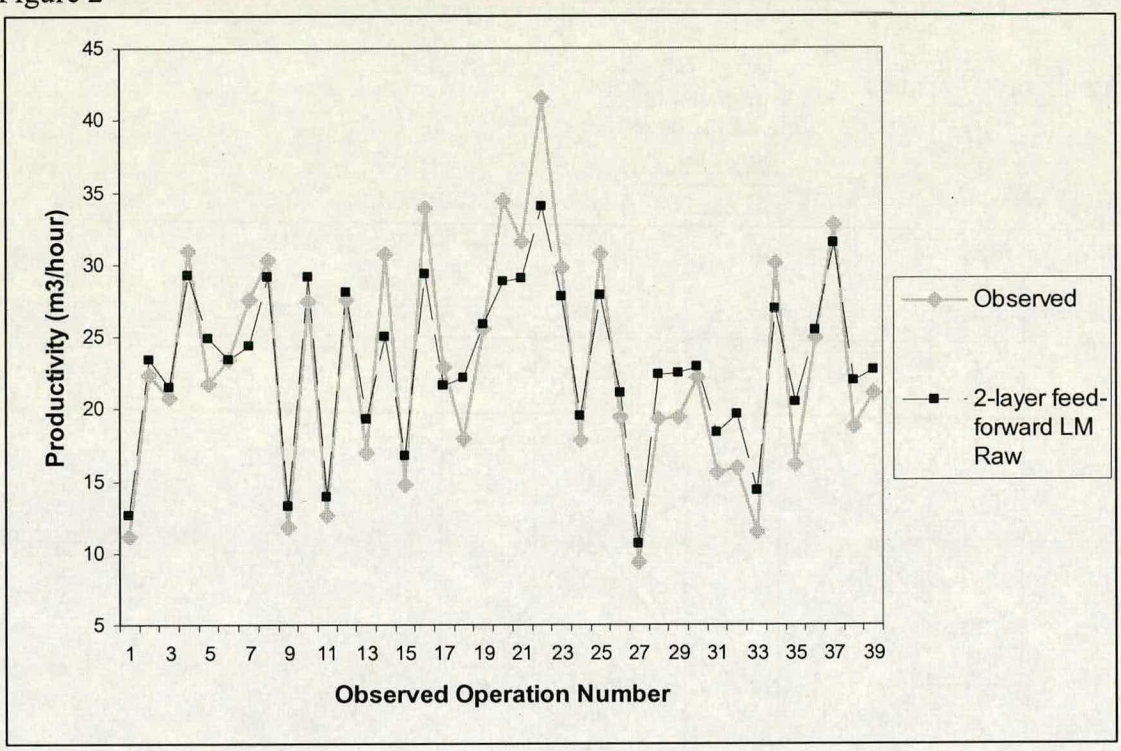


Figure 3

Collected Variables	Output
Month of Operation	
Type of Operation	
Truck Volume (m^3)	
Total Operation Volume (m^3)	
Average Interarrival Time (minutes)	Productivity (m^3/hour)
Number of Loads in Operation	
No. of Accepted Loads	
No. of Rejected Loads	

Table 1

Table 2

	Network Type															
	Feed-Forward								Elman							
No. Layers	2	2	2	2	3	3	3	3	2	2	2	2	3	3	3	3
Training Algorithm	LM	LM	SCG	SCG	LM	LM	SCG	SCG	LM	LM	SCG	SCG	LM	LM	SCG	SCG
Data Type	Raw	Scaled	Raw	Scaled	Raw	Scaled	Raw	Scaled	Raw	Scaled	Raw	Scaled	Raw	Scaled	Raw	Scaled
No. of Times Developed	4	4	4	4	2	2	4	4	4	4	4	4	4	4	4	4
Average Test MSE (m^3/hour)	10.76	22.28	11.32	22.57	10.60	22.21	9.65	22.32	32.87	22.62	12.06	22.32	29.99	22.31	9.14	22.41
Standard Deviation Test MSE (m^3/hour)	2.99	0.01	0.63	0.01	4.95	0.00	2.10	0.01	8.00	0.00	0.81	0.01	4.56	0.00	2.14	0.00
No. Neurons (in hidden layers)	10	2	18	4	12-6	4-22	20-20	6-18	12	12	6	6	6-2	6-2	22-16	20-12
Activation Functions between layers	L-P	L-T	P-P	P-P	L-P-P	T-P-P	T-P-P	L-T-T	P-P	T-P	P-P	L-T	P-P-P	P-L-T	T-P-P	L-T-T

Table 3

Goodness of Fit Test	NN Model					
	2-Layer Feed-forward (LM Trained)	2-Layer Feed-forward (SCG Trained)	3-Layer Feed-forward (LM Trained)	3-Layer Feed-forward (SCG Trained)	2-Layer Elman (SCG Trained)	3-Layer Elman (SCG Trained)
K-S D-Value	4 =	3	2	4=	1	6
Theil, U	2	1	3	4	6	5
RMS Proporional	1	4	2	3	5 =	5 =
R^2	1	4	2	3	5	6
$+/- 2\sigma$	1 =	1 =	1 =	5 =	1 =	5 =
Totals	9	13	10	19	18	27
Overall Ranking	1st	3rd	2nd	5th	4th	6th

Figure and Table Captions

Figure 1a: Feed-forward NN (adapted from The MathWorks, 2003)

Figure 1b: Elman recurrent NN (adapted from The MathWorks, 2003)

Figure 2: Residual plot for linear regression of collected data against productivity

Figure 3: A plot of the correlated inspection results for the 2-layer feed-forward (LM trained) network

Table 1: The collected variables to be considered in this study

Table 2: Training and testing results for the NN models

Table 3: Ranked correlated inspection results

In-Situ Concreting: Identification and Quantification of Contractor Costs

Darren Graham[1], Simon D. Smith[2] and Iris D. Tommelein[3]

[1] Postgraduate Student, School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, United Kingdom.

[2] Lecturer in Construction and Project Management, School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, United Kingdom. Tel: +44 131 650 7159; Fax: +44 131 650 6781.

simon.smith@ed.ac.uk [*Corresponding Author*]

[3] Professor of Engineering and Project Management, Department of Civil and Environmental Engineering, The University of California at Berkeley, Berkeley, CA 94720-1710, United States of America.

For: Construction Management and Economics

Abstract

In-situ concreting commonly involves using ready-mixed concrete supplied by an external manufacturer and the contractor's cost for this may be significant. Previous research to minimise this cost has focussed upon optimising the performance of concreting operations, through the optimal allocation of resources. This appears to have been done without consideration of other costs involved in such operations. The aim of this research is to determine the costs that should be reduced to minimise the contractor cost of in-situ concreting operations and identify method(s) of achieving this minimisation.

The costs and the contributor(s) to these costs were identified through consultation with industrial practitioners. A quantitative analysis was performed on a data set collected from UK construction projects to determine the significance of each cost and the contributor(s) to each cost. A comparison of the identified costs with those from existing sources indicated that the results are generally applicable in both the UK and the USA.

The significant costs were found to be: the RMC material, plant hire costs, idleness of plant and labour (placing team) and truck mixer surcharges (for trucks delayed on site by contractors). These costs should be reduced.

A systemic cost reduction would work to reduce the contractor's cost of in-situ concreting by including all of the significant costs, identified above, in any solution. Such a systemic solution would require a fundamental re-management of the way that in-situ concreting is performed today. This could involve a reconfiguration of the resource supply chains, coupled with production control methods, such as more sophisticated planning using simulation.

Key Words: ready-mixed concrete; cost; quantitative analysis; materials management; production management

Introduction

Global ready-mixed concrete (RMC) production is significantly increasing. In the USA it has more than doubled over the past 30 years (NRMCA, 2005); while production value has increased from: €1.24 billion to €1.42 billion in UK; €1.9 billion to €2.68 billion in France; and €780 million to €3.62 billion in Spain over the period 1995 - 2003 (Eurostat, 2005).

As an important construction material, RMC should be manufactured, transported and placed with a high level of efficiency and at a minimum cost to the supplier, contractor and ultimately the client. Unfortunately this appears not to be the case; previous studies such as Halpin and Woodhead (1976), Camillo (1996), Smith (1998, 1999), Zayed and Halpin (2001) and Dunlop and Smith (2003) have found that the stochastic nature inherent in the RMC supply and placing system (RMC system) makes it difficult to plan and manage, resulting in increased pour durations; poor utilisation of both the truck mixers (that transport the RMC) and the project resources (concrete pump, or crane and skip, labour, etc.); and an increased cost of construction.

Previous research projects have developed numerous deterministic models (for example, Dunlop and Smith, 2004) and stochastic models (for example, Sawhney et al., 1999) of the RMC system in a bid to make better predictions of the achievable performance (production rate, cost, etc.), given different allocations of the resources involved. Some of these have proved to be capable of accurately estimating the production rates of the system, for instance Graham et al. (2004) presented a model *CBRSim* that could estimate to within +/- 3% accuracy. Additionally, a number of

previous studies have suggested that a minimisation of cost can be achieved through optimising the allocation of resources, given the operating conditions (for example, Zayed and Halpin, 2001).

While these modelling and optimisation methods could result in a decrease in the cost of RMC operations it is felt that this reductionist view of the RMC system may not identify all of the costs that comprise the system. Furthermore, it could be that the contractor's resource cost may not be the most significant, and that the above studies have focused upon elements of the system that may not have the most impact on a contractor's profitability.

It is clear from the dates of the above publications that this is not a recent problem. However, as stated above, the construction industry is placing an increasing reliance on RMC as a construction material and this trend is the catalyst for this research work. If, in a climate of increasing use of the material, RMC continues to be placed with little attention being paid to inefficiencies, this could increasingly impede on the profitability and competitiveness of construction contractors and, consequently, the value provided to project clients.

Research aims and methodology

The aims of this study are to define the sources of contractor cost and outline methods for their reduction, in the case where RMC is supplied externally. Using data collected from real construction projects an investigation was therefore undertaken to:

- Identify and quantify the costs that exist in the RMC system for a contractor;

- Expose the sources of these costs and to assess the significance of each source.
- Define areas of focus for contractor cost reduction strategies.

This article presents this investigation and covers the following stages:

- A presentation of the sources of the project data.
- A definition of the RMC system.
- The costs associated with this system for contractors are defined, based upon discussions with contractors and batch plant managers.
- Contractor costs in the system are quantified
- The contributing factors to these costs are examined.
- To investigate the applicability of the results to both the UK and the US markets, a comparison is made between the results of this research and the cost information provided: by Spon's Civil Engineering and Highway Works Price Book 2005 (Spon's, 2005) for the UK; by RS Means Building Construction Cost Data 1998 (RS Means, 1998) for the USA.
- Areas of focus for contractor cost reduction strategies are defined.
- Finally, conclusions are drawn from the results of this research project.

Data Sources and Details

The data to be analysed in this study was collected through observations of two construction projects in Scotland, UK.

Project one, an £80 million Private Finance Initiative contract, was located in the Northeast of the country and involved the construction of four separated water

treatment facilities and was undertaken by Contractor A. There were two suppliers (A and B) of RMC involved in this project – located at 5 and 18 miles from the project site. A total of 184 concreting operations (pours) and 14,900 m^3 of placed concrete were observed.

Project two was located in the East of Scotland, and was financed with a Private Finance Initiative (PFI) contract worth £90 million. This project involved the construction of two wastewater treatment plants by Contractor B. Two suppliers (C and D) provided the RMC for this project – located at 8 and 16.5 miles from the project site. 48 concreting operations (pours) and a total volume of placed concrete of 6,200 m^3 was observed.

For generic conclusions to be drawn from this study, it is important to ensure that the project data is evenly and well-distributed. In past RMC studies a measure of how well a data set is distributed has been made by presenting a histogram of: observed pour volumes (e.g. Anson et al., 2002); and productivities (Dunlop and Smith, 2003) – two key variables of the system. A histogram of the pour volumes and line-graph of the productivities in the project data is shown in figure 1.

The histogram shown in figure 1 indicates a higher proportion of pours of 20 m^3 to 150 m^3 than those above or below this volume. When the average productivity of each of the volume categories is plotted (as in figure 1) a significant difference is shown between the average productivities of 80 m^3 or less pours and those above this value; pours with a volume more than 80 m^3 have a higher rate of productivity. Thus, the data set could reasonably be thought of as two distinct groups, one consisting of

pours 80 m^3 or less (small pours) and another of pours greater than 80 m^3 (large pours) and these will be the two groups considered in this study. As each of these groups is more evenly distributed than the original project data it is concluded that the results of the analysis are generic and applicable to a wide range of RMC operations.

FIGURE 1

The data variables collected from observations are shown in Table 1.

TABLE 1

The data was recorded in four worksheets in a computer spreadsheet, an example of the recorded times are shown in figure 2; the remaining information is held in separate worksheets not shown.

FIGURE 2

Definition of RMC Supply and Placing System (RMC System)

The RMC supply and placing system was defined following numerous visits to construction projects and RMC batching plants in both UK and USA and is shown in figure 3. The system can be separated into two processes: supply and placing. The RMC supply process is cyclical and starts and ends at the batching plant. A truck mixer, when it arrives at the site and gets to the front of the queue at the pump (if

any) is then tested for quality. If the quality is acceptable the truck mixer enters the RMC placing process – and this can be seen in a separate part of Figure 3.:

FIGURE 3

In attempting to simplify this model of the process, and to concentrate on its pertinent components, the following have been ignored:

- The supply chain of the raw materials of RMC (e.g. cement)
- The supply chains for the reinforcement steel, formwork and any scaffolding.
- The processes of erecting/dismantling of scaffolding, formwork and fixing reinforcement steel.
- Any need for the driver of a truck mixer to exit the vehicle has been ignored. This commonly occurs at both the batching plant and site, and is assumed to have been absorbed in the queues 1 and 2 in the above definition.
- Washout of truck mixers at the batching plant, principally because it is not performed consistently and is therefore difficult to model.

Identification of Costs in the RMC System

With consideration of the system definition provided in figure 3 and the boundaries indicated by those components which have been ignored, the authors collaborated with several managers from RMC batching plants (in both USA and UK) and construction contractors (in UK) to identify the main costs in the system. These costs are presented in an Ishikawa (Fishbone) diagram in figure 4. Note that indirect costs (including storage costs) have been ignored.

FIGURE 4

The costs for a contractor are as follows:

- **Plant-hire costs**. The plant required to perform RMC operations can differ , depending on particular operational characteristics. However, standardised plant across RMC operations was observed, including a truck mounted concrete pump, vibrator (poker) units, and mechanical trowels. Other pieces of equipment have been included in the study and are accounted for in the quantitative analysis of costs presented in the next section.

  Whether or not the plant is owned outright by the contractor it is often the case that plant is hired (rented) and thus the contractor faces plant hire costs, which are usually charged on a fixed daily basis.

- **Plant-operating costs.** The operating costs of the plant are partially a function of pour duration, fuel prices, plant running during idle periods in the pour, overall fuel consumption and repair costs. Whilst fuel costs may vary between markets, contractors should be be concerned about fuel being wasted during idle spells in a pour, where the plant remains running.

- **Labour costs (concrete placing team).** RMC operations are performed in teams, often consisting of a ganger (or foreman), a concrete pump operator, concrete placing personnel, general labourers and concrete finishers. The

labour force is usually paid an hourly rate, the overall cost therefore being proportional to duration.

- **Truck mixer surcharge.** The RMC supplier, often operating on tight margins, will attempt to penalise unnecessary truck mixer idleness. They therefore usually have a tolerance for the length of time that a truck mixer may be used on a site to discharge its load. This tolerance was found to be generally 5 minutes per cubic metre of RMC being placed. Beyond the tolerance the RMC supplier attaches a surcharge of RMC of roughly £2 per minute over the tolerance.

- **Cost of RMC material**. The final contractor cost identified in the RMC system for the material. The quoted price of RMC (provided by the supplier) is often inclusive of delivery charges – as was the case in the studies projects – and can vary significantly according to the design mix of the concrete. Typically, the unit price of concrete is in the range of £40/m^3 to £75/m^3 (Spon's, 2005).

Quantification of Contractor Costs in the RMC System

The identified costs can be quantified for each of the observed RMC operations in the project data. These quantities were calculated using the following formulae:

$$RMC\ Material\ Costs\ (£) = Unit\ Cost\ of\ RMC\ (£/m^3) \times Pour\ Volume\ (m^3) \quad (1)$$

$$Plant\ Hire\ Costs\ (£) = Daily\ Hire\ Rate\ (£/day) \times No.of\ Days\ for\ Operation \quad (2)$$

$$Plant\ Operating\ Costs\ (£) = Unit\ Operating\ Cost\ (£/hour) \times Operation\ Duration\ (hours) \quad (3)$$

$$\textit{Labour Hire Costs}\,(£) = \textit{Duration}\,(\textit{hours}) \times \left[\textit{No. Persons in Each Job}\,(\textit{e.g. labourer}) \times \textit{Unit Cost of Each Job}\,(£ / \textit{hour})\right] \tag{4}$$

$$\textit{Truck Mixer Surcharge}(£) = \begin{cases} [\text{Truck Mixer Time at Project (minutes)} - \text{Tolerance Threshold (minutes)}] * £\ 2/\text{minute} & \textit{If Truck Mixer Time at Project} > \textit{Tolerance Threshold} \\ 0 & \textit{If Truck Mixer Time at Project} <= \textit{Tolerance Threshold} \end{cases} \tag{5}$$

$$\textit{Contractor Cost}\,(£) = \textit{RMC Material Costs} + \textit{Plant Hire Costs} + \textit{Plant Operating Costs} + \ldots\ldots + \textit{Labour Hire Costs} + \textit{Truck Mixer Surch}\,\text{arg}\,e \tag{6}$$

The values for pour volume, duration of pour and the make-up of the labour force were recorded for each pour in the project data set. The unit costs and hire rates were obtained from RMC suppliers and construction contractors and are quoted at 2005 prices. The above calculations are illustrated in the following example.

The relevant details of the RMC operation for this example are presented in table 2. Entering these details into equations (1) to (6) results in:

RMC Material Costs = 35 * 42 = £1470.00 (1)

Plant Hire Costs = 345 * 1 = £345.00 (2)

Plant Operating Costs = 20 * 3.72 = £ 74.40 (3)

Labour Hire Costs = 3.72 * [1 (ganger) * 15.00 + 5 (concrete placers) * 10.00 + 2 (labourers) * 9.50 + 1 (plant operator) * 14.00] = £ 364.56 (4)

Truck Mixer Surcharge = £ 0 (Time at Project < Tolerance) (5)

Contractor Cost = £ 1470.00 +345.00 + 74.40 + 364.56 + 0.00 = £ 2253.96 (6)

TABLE 2

A total of 168 RMC material costs, plant hire costs, plant operating costs, labour hire costs and truck mixer surcharges were calculated for small pours; and 68 for large

pours. The total worth of each of the mentioned costs in each project data set was calculated, and the results for small pours are presented in figure 5.

FIGURE 5

The total contractor cost for small pours was £543,000. Of this total the majority (73%) can be attributed to RMC material cost. 25% was due to plant and labour costs, while the remaining 2% was cost due to truck mixer surcharges.

Similar results were found for large pours. Notably the cost of RMC material is even more important in pours of this size, as 86% of the contractor cost (of £671,000) could be attributed to the acquisition of RMC.

The results of this quantitative analysis provide justification for this study. This is because previous studies have centred upon reducing RMC costs through reducing the cost of plant and labour by optimally allocating resources. While this remains a useful strategy, it ignores the significance of the cost of RMC. To reduce the cost of RMC by only 10% would have resulted in savings of £54,320 (in small pours) and £67,120 (in large pours) for the contractors in the observed projects.

The above notwithstanding the potential savings in plant and labour costs should not be neglected, as together they comprise 25% of contractor cost (in small pours) and 12% (in large pours). Additionally, the truck mixer surcharge appears to be an unnecessary expense that could be removed through better production control.

Taking the above quantitative analysis into account it is worthwhile examining the contributors to the contractor costs in the RMC system to aid the identification of areas of focus for cost reduction strategies.

Contributors to Contractor Costs in the RMC System

Identified through the observation of construction projects and consultation with construction contractors and RMC batch plants, in both the UK and the USA, these contributors were subsequently quantified through an analysis of real construction data. This quantification is presented in figure 5 for small pours (80 m^3 or less); similar results were found for large pours (> 80 m^3). The contributors to each cost in the RMC system shall now be discussed:

Cost of RMC material. The contributors to RMC material costs are as follows:

- Raw materials. That is, cement, fine aggregates (sand), coarse aggregates, water and either mineral or chemical admixtures. The price of each of these raw materials is governed by market forces, and an estimate of these prices was provided by two RMC batch plants. It was found in this research project that raw materials account for 54% of the total contractor cost in small pours; a similar result was found for large.
- Transportation. There are a number of factors in the transportation of batched concrete, for example fuel prices; truck mixer capital costs, maintenance costs and operating costs; periods of truck mixer idle time; and operator costs. An estimate of the unit cost of transportation was obtained from two RMC batch plants and used in the quantitative analysis

performed in this study. Transportation was estimated to account for 5% of the total contractor costs in small and large pours.

- Supplier overhead. Batch plant capital costs, operating costs; maintenance costs; labour costs; business rates; and utility charges. Such overheads were estimated to contribute 12% of the total contractor cost in the studied small pours; 10% in large pours.
- Supplier profit. Commonly quite small: a survey of the RMC batch plants visited in this project revealed that on average the supplier's profits were equal to just 2% of the total contractor cost in small pours; rising to 4% of the total contractor cost in large pours.

Plant hire and operating costs. The contributors to plant hire and operating costs are as follows:

- Idle plant (outwith pour). Plant is usually hired at a fixed daily rate. In this study it was assumed that a standard day, on a construction project, is of 12 hours duration. It was found that the plant is often not used for significant periods of the 12 hour period at a cost to the contractor. This contributes an estimate of 3% of the total contractor costs in small pours and 1% in large pours.
- Set up costs. It has been estimated by the contractors consulted in this study that setting up plant could contribute 1% of the total contractor cost in a project.
- Idle plant (operating cost). When there is a gap in the RMC supply the plant is rarely switched off, and a contractor must pay for this. This quantity has been calculated to be worth 1% of the total contractor cost.

- Plant hire (value-adding). This is the hire cost that can be attributed to the plant being used to place RMC and thus could be called a 'value-adding' cost. This has been estimated to equate to 5% of the total contractor cost in small pours.
- Plant operating cost (value-adding). Similar to the previous contributor to cost, this is essential to achieving the overall goal of the RMC system. However, it may be possible to minimise this cost, while maintaining the production rate of the system using some form of production control. Plant operating cost contributes 2% to the total contractor cost.

Labour hire cost. This can be separated into two groups of factors: those where labour is working on the placing of RMC (value-adding); those where labour is standing idle. For small pours, the former account for 11% of the total contractor costs, while the latter is worth 2%.

Three reasons for labour standing idle were identified in the observation of RMC pours:

- There were an insufficient number of truck mixers being used to deliver RMC resulting in gaps in the supply of the material;
- When a load of RMC is rejected, most commonly due to the material failing a slump test, a gap in the supply of RMC is created;
- When the pump suffered a breakdown.

Quantifying the costs associated with these three reasons, it was found that rejected loads and pump breakdowns occurred too infrequently to have anything other than a negligible affect on contractor cost..

It was found that 2% of the total contractor cost could be attributed to having an insufficient number of truck mixers in the RMC system.

Truck mixer surcharge. There a number of reasons for the supplier's idleness tolerance being exceeded, including: pump breakdown; pump operator skill level; truck mixer driver skill level; accidents on the project site; poor site access routes; site unprepared in terms of labour availability, reinforcement bars not in place or formwork not in place; excessive numbers of truck mixers in RMC system resulting in queuing on the project site.

Comparison with Existing Construction Cost Information

It is important to ensure that the results of the quantitative analysis can be applied to different types of projects from those that the project data came from. Therefore, a comparison was made between the results of the quantitative analysis performed in this research work and existing construction cost information produced by Spon's (in the UK) and RS Means (in the USA). This comparison should provide some evidence of the applicability of the results in both the UK and the USA, thereby expanding the number of beneficiaries.

RS Means and Spon's provide estimates of the breakdown of contractor costs in performing RMC operatings (RS Means, 1998; Spon's, 2005), and these estimates, for small pours, are shown in table 3.

It can be seen that the cost breakdown from this study is similar to that of RS Means and Spon's, which indicates that the results of this study are generally applicable to both the UK and the USA construction industries. However, there are differences of note:

- RS Means do not recognise the impact of truck mixer surcharge on contractor costs; a cost found to be significant in this study. Nevertheless, it was confirmed through consultation with construction contractors and RMC suppliers in the USA that a truck mixer surcharge is a factor of contractor cost in the RMC system.

- The RS Means edition was published in 1998 and is based in the USA, whereas the data used in this study and Spon's cost estimates are at 2005 prices and taken from projects performed in the UK.

  Since 1998 the cost of raw materials has increased globally (DTi, 2004). In the UK this has particularly been the case due to the introduction of the UK Government's Aggregates Levy in 2002 (DEFRA, 2005). This may account for the difference (of 1%) between the findings of this study (and Spon's) and the estimates of raw material cost provided by RS Means.

The price of crude oil has risen significantly from 1998 to 2005, resulting in a substantial increase in on-site plant operating costs. This increase is reflected in the differences between the quantitative analysis (and Spon's) and the estimates from RS Means.

- There is a significant difference between the estimates of labour hire provided by all three sources, possibly reflecting the differences in the labour markets, and hence labour costs, in the UK in general (i.e. in Spon's rates), in Scotland (where the projects were located) and the USA.

TABLE 3

Areas of Focus for Contractor Cost Reduction Strategies

A number of areas that contractor cost reduction strategies could focus upon can be defined from the results of this study. They are as follows:

1) **Cost of RMC material.** Represents the most significant cost for contractors in RMC operations. In the studied projects, RMC materials comprised 73% of small pour costs, and 86% of large pour costs. This represented a total cost of £973,700 in the studied projects.

   The inherent nature of three of the contributors to this cost, identified in the quantitative analysis of the project data, indicates that the cost of RMC may be reduced. These contributors are transportation (5% of total cost), supplier's overhead (12% total cost) and supplier's profit (2% total cost). If the supply chain were vertically integrated, by removing the external RMC

supplier, it may be possible that some of the costs could be removed for the contractors. For example, the removal of an external supplier of RMC would immediately provide a cost saving of 2% of total cost to the contractor that was previously being taken by the RMC supplier as profit; RMC production would not be a profit making enterprise for the contractor – merely a method of cost-saving.

There are two types of truck mixer available for a contractor to use as an alternative to procuring RMC from a batch plant external to the project site (Sonnenberg, 1998; Ferraris, 2001), and thereby implementing such a vertical integration of the RMC supply chain: mixers that produce batches of RMC; and continuous or volumetric mixers. An analysis of the practical introduction of these technologies in the RMC system shall be the subject of future research.

2) **Cost of hiring plant.** A significant amount of cost (3% of total) is incurred through plant sitting idle outside of working hours as operations rarely last a full-working day, while the plant is paid for at a flat daily-rate. The cost of hiring plant comprises 24% of the plant and labour costs in small pours and 17% in large pours, at a cost to the contractor of £45,700 in the studied projects.

However, there is usually a sensible, practical reason why pours cannot be performed consecutively, in a manner that maximises the duration of use of the hired plant. For example, there could be a need to erect formwork or steel

reinforcement prior to the next concrete pour and this would result in a pause in the concreting operations. Thus, this issue cannot be readily resolved under the common system of hiring plant at a flat daily-rate.

Rather, in some instances it may prove worthwhile for a contractor to re-negotiate plant-hire contracts to an hourly-based charging system. This would be most beneficial if the project site is located close to the plant supplier.

This above strategy is not without its pitfalls. The savings made by this strategy are likely to be off-set, to a degree, by an increase in costs due to additional set-up times (which currently constitute 1% of total contractor cost in small and large pours) and transportation durations of the plant to the project site.

An alternative to the above strategy would be for a contractor to purchase the plant, given suitable market conditions (e.g. sufficient future demand for the contractor's services). This is effectively a vertical integration of the supply chain, because the plant supplier is being removed on a project-by-project basis, as the plant would be transferable between projects and pours.

3) **Cost of plant and labour standing idle and truck mixer surcharge.** A balance should be achieved between having too few and too many truck mixers to deliver RMC in the system. Too few results in plant and labour waiting for RMC deliveries (3% of total contractor cost); too many resulting

in truck mixers queuing on site. The consequences of this are non-value adding costs, in both truck mixer idleness and supplier surcharges.

A systemic solution, which minimises as many cost contributors as possible, should be sought. This has been hinted at already, via vertical integration of the resources (materials, plant, etc.) supply chains (Hill and Moore, 2004), coupled with a system of production control, such as process simulation, to achieve a balance between the utilisation of truck mixers and labour, while maximising the productivity of the entire system.

Vertical integration of the supply chains of the resources could entail:

1. The contractor producing its own RMC, and thereby removing an external RMC supplier from the system;
2. The contractor purchasing its own plant for use in concrete placing, and thereby removing the plant supplier from the system.

Conclusions

In future attempts to reduce the cost of in-situ concreting operations for contractors a focus should be placed upon a number of factors, rather than improving the performance of these operations through the optimum allocation of resources. These factors are: cost of the RMC material; cost of hiring plant; cost of plant and labour standing idle during the operation; and, cost of surcharges added for truck mixers being retained on a project site for excessive periods of time.

To this end, this study has proposed that a systemic solution should be adopted to minimise the cost of RMC operations for contractors, which incorporates measures that aim to reduce each of the factors outlined above.

This systemic solution involves a combination of production control measures, such as better planning of the operation using simulation methods, and a re-management of the supply chains of the resources involved in the RMC system.

A vertically integrated supply chain would reduce the cost for contractors by: removing the supplier's profit margin; removing any surcharge for the use of truck mixers in delivering RMC; reducing the supplier's overhead – as RMC production would be a much smaller enterprise for a contractor supplying only its own projects; reducing the amount of money spent on transportation of the material to the project site. It is likely that, through negotiation with the raw materials suppliers, the cost of delivering raw materials could be minimised. This, coupled with the production of RMC on (or close to) the project site, could result in a significant reduction in the cost of transporting RMC.

To practically implement such a systemic solution, as that proposed above, will require an examination of how practical it would be for a contractor to produce RMC and purchase plant. This is to be the subject of future investigations.

References

Anson, M., Tang, S. L. and Ying, K. C. (2002). Measurement of the performance of ready mixed concreting resources as data for system simulation. *Construction Management and Economics*, 20 (3), 237 – 250.

Camillo, J. (1996). Controlling the flow of truck mixers. *Concrete J.*, 14 (4), 250 – 256.

Department for Environment, Food and Rural Affairs (DEFRA). (2005). *Recycling and waste: aggregates levy sustainability fund in England.* http://defra.gov.uk/environment/waste (Accessed: 26/07/05).

Department of Trade and Industry (DTi). (2004). *Construction statistics annual 2004*. HMSO, London., UK.

Dunlop, P. and Smith, S. D. (2003). Estimating key characteristics of the concrete delivery and placement process using linear regression analysis. *Civil Engineering and Environmental Systems*, 20 (4), 273 – 290.

Dunlop, P. and Smith, S. D. (2004). Planning, estimating and productivity in the lean concrete pour. *Engineering, Construction and Architectural and Management*, 11 (1), 55 – 64.

Eurostat (2005). DS-008451-PRODCOM-ANNUAL-2663100-Ready mixed concrete. http://epp.eurostat.cec.eu.int. (Accessed: 04/05/05).

Ferraris, C. F. (2001). Concrete mixing methods and concrete mixers: state of the art. *J. Res. Natl. Inst. Stand. Technol.*, 106 (2), 391 – 399.

Graham, D., Smith, S. D. and Crapper, M. (2004). Improving construction simulation with a case based reasoning input. *Civil Engineering and Environmental Systems*, 21 (2), 137 – 150.

Halpin, D. W. and Woodhead, R. W. (1976). *Design of construction and process operations*. John Wiley and Sons, Inc., NY., USA.

Hill, W. and Moore, B. (2004). Business models in a consolidating ready mixed concrete industry. *In: Concrete in Focus*. National Ready Mixed Concrete Association (NRMCA), Silver Spring, MD., USA.

National Ready Mixed Concrete Association (NRMCA). (2005). Ready mixed concrete production statistics. http://www.nrmca.org (Accessed: 04/07/05).

RS Means. (1998). *Building construction cost data*. 56th edition. RS Means Co., Inc., Kingston, MA., USA.

Sawhney, A., Abudayyeh, O. and Chaitavatputtiporn, T. (1999). Modeling and analysis of concrete production plant using Petri nets. *J. Computing in Civil Engineering, ASCE*, 13 (3), 178 – 186.

Smith, S. D. (1998). Concrete placing analysis using discrete event simulation. *ICE Structures and Buildings*, 128 (4), 351 – 358.

Smith, S. D. (1999). Modelling and experimentation of the concrete supply and delivery process. *Civil Engineering and Environmental Systems*, 16 (2), 93 – 114.

Sonnenberg, R. (1998). Concrete mixers and mix systems. *Conc. Precast Plant Technol.*, 64, 88 – 98.

Spon's. (2004). *Spon's civil engineering and highway works price book 2005*, Langdon, D. (ed), 19th Edn. Spon Press, London, UK.

Zayed, T. M. and Halpin, D. W. (2001). Simulation of concrete batch plant production. *J. Construction Engineering and Management, ASCE*, 127 (2), 132 – 141.

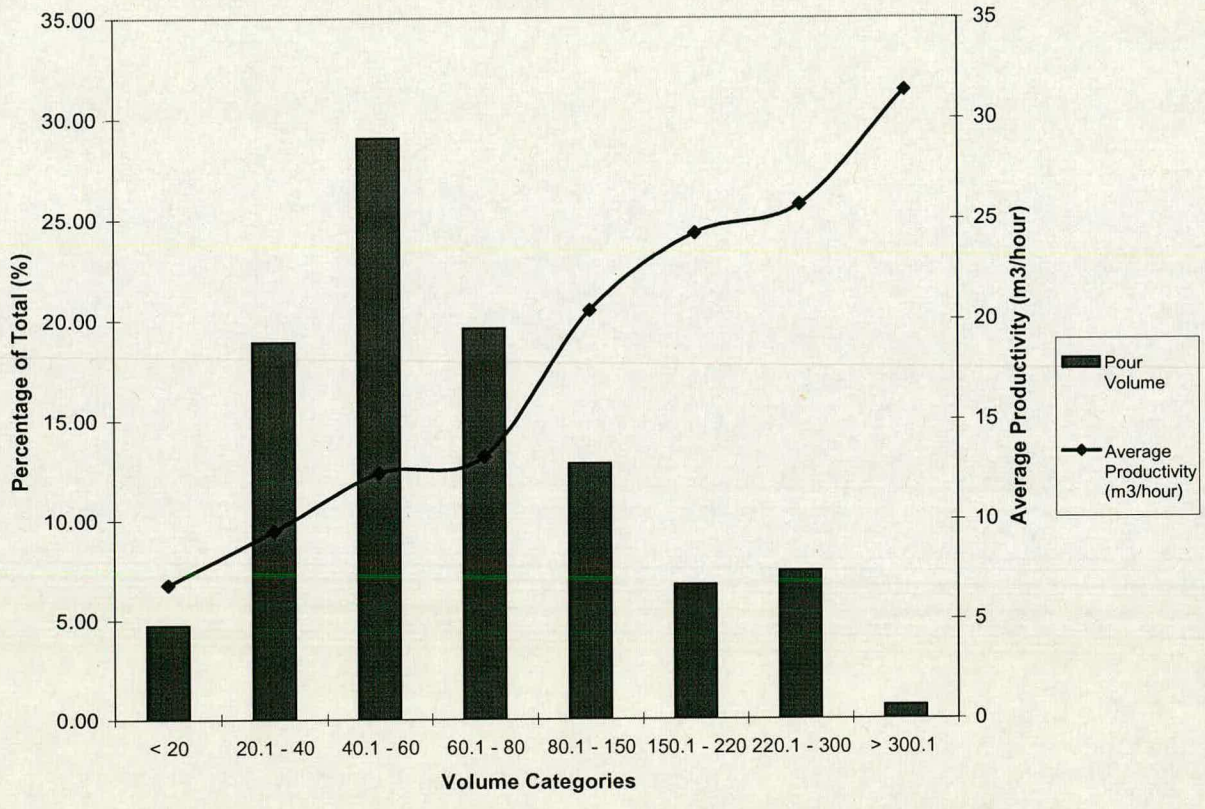

35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
Percentage of Total (%)
35
30
25
20
15
10
5
0
Average Productivity (m3/hour)
< 20
20.1 - 40
40.1 - 60
60.1 - 80
80.1 - 150
150.1 - 220
220.1 - 300
> 300.1
Volume Categories
Pour Volume
Average Productivity (m3/hour)

Figure 1

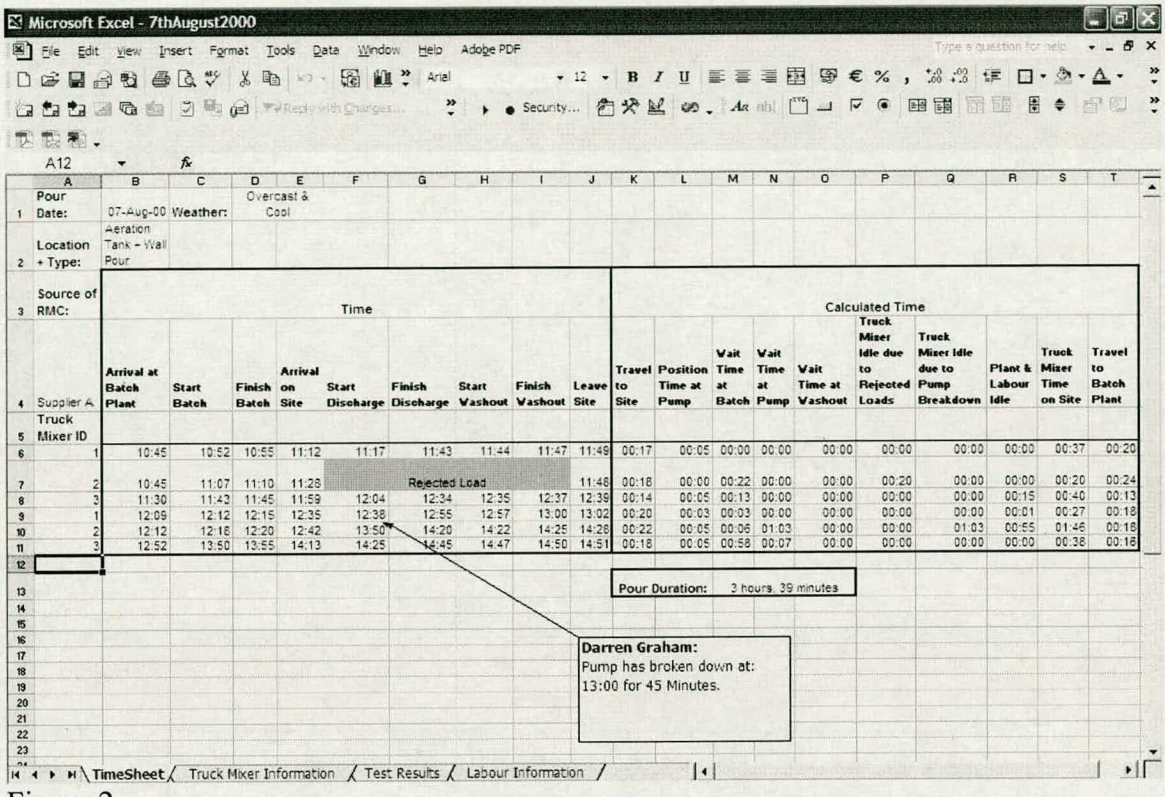

Microsoft Excel - 7thAugust2000
File Edit View Insert Format Tools Data Window Help Adobe PDF
Pour Date: 07-Aug-00 Weather: Overcast & Cool
Location + Type: Aeration Tank - Wall Pour
Source of RMC:
Time
Calculated Time
Supplier A
Truck Mixer ID
Arrival at Batch Plant
Start Batch
Finish Batch
Arrival on Site
Start Discharge
Finish Discharge
Start Washout
Finish Washout
Leave Site
Travel to Site
Position Time at Pump
Wait Time at Batch
Wait Time at Pump
Wait Time at Washout
Truck Mixer Idle due to Rejected Loads
Truck Mixer Idle due to Pump Breakdown
Plant & Labour Idle
Truck Mixer Time on Site
Travel to Batch Plant
1 10:45 10:52 10:55 11:12 11:17 11:43 11:44 11:47 11:49 00:17 00:05 00:00 00:00 00:00 00:00 00:00 00:00 00:37 00:20
2 10:45 11:07 11:10 11:28 Rejected Load 11:46 00:18 00:00 00:22 00:00 00:00 00:20 00:00 00:00 00:20 00:24
3 11:30 11:42 11:45 11:59 12:04 12:34 12:35 12:37 12:39 00:14 00:05 00:13 00:00 00:00 00:00 00:00 00:15 00:40 00:13
1 12:09 12:12 12:15 12:35 12:38 12:55 12:57 13:00 13:02 00:20 00:03 00:03 00:00 00:00 00:00 00:00 00:01 00:27 00:18
2 12:12 12:18 12:20 12:42 13:50 14:20 14:22 14:25 14:26 00:22 00:05 00:06 01:03 00:00 00:00 01:03 00:55 01:46 00:18
3 12:52 13:50 13:55 14:13 14:25 14:45 14:47 14:50 14:51 00:18 00:05 00:58 00:07 00:00 00:00 00:00 00:00 00:38 00:16
Pour Duration: 3 hours, 39 minutes
Darren Graham:
Pump has broken down at:
13:00 for 45 Minutes.
TimeSheet Truck Mixer Information Test Results Labour Information

Figure 2

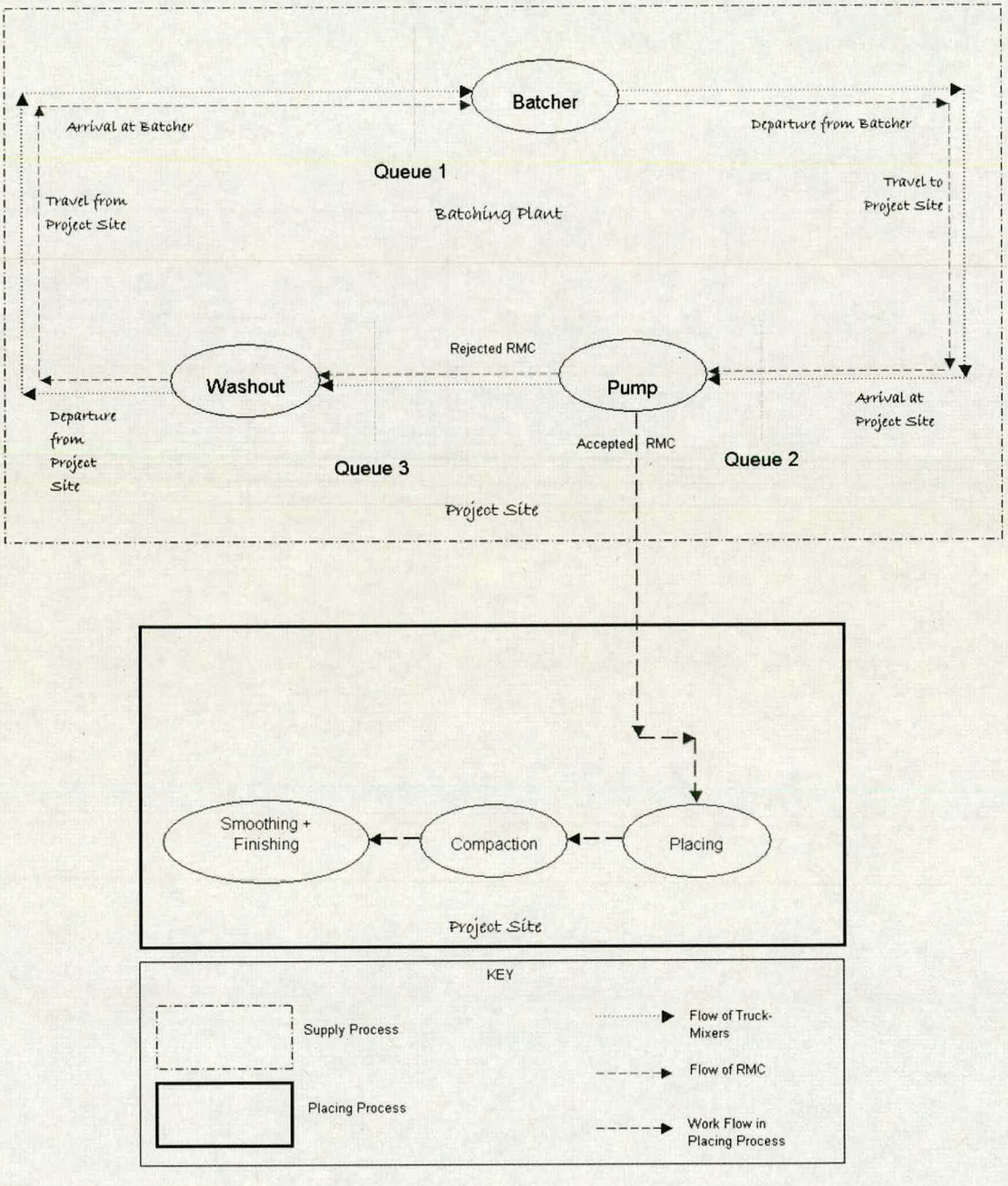

Batcher
Arrival at Batcher
Departure from Batcher
Queue 1
Travel from Project Site
Batching Plant
Travel to Project Site
Rejected RMC
Washout
Pump
Departure from Project Site
Arrival at Project Site
Accepted RMC
Queue 3
Queue 2
Project Site
Smoothing + Finishing
Compaction
Placing
Project Site
KEY
Supply Process
Placing Process
Flow of Truck-Mixers
Flow of RMC
Work Flow in Placing Process

Figure 3

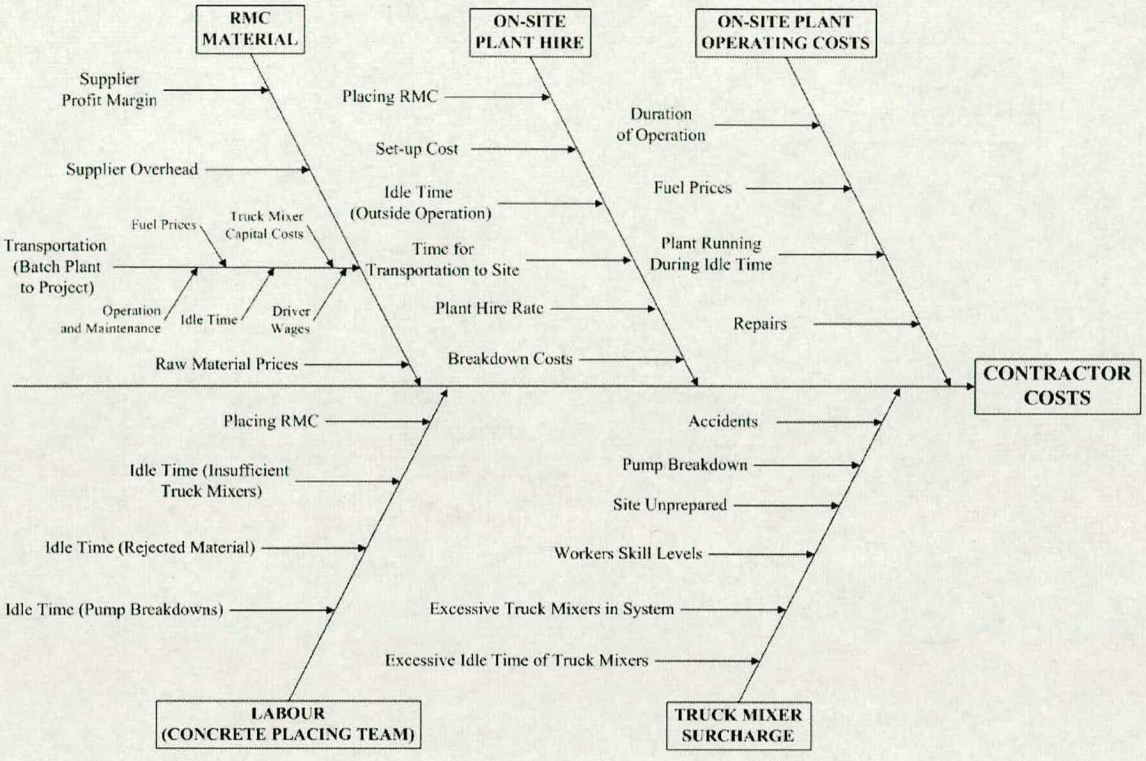

RMC MATERIAL
ON-SITE PLANT HIRE
ON-SITE PLANT OPERATING COSTS
Supplier Profit Margin
Supplier Overhead
Transportation (Batch Plant to Project)
Fuel Prices
Truck Mixer Capital Costs
Operation and Maintenance
Idle Time
Driver Wages
Raw Material Prices
Placing RMC
Set-up Cost
Idle Time (Outside Operation)
Time for Transportation to Site
Plant Hire Rate
Breakdown Costs
Duration of Operation
Fuel Prices
Plant Running During Idle Time
Repairs
CONTRACTOR COSTS
Placing RMC
Idle Time (Insufficient Truck Mixers)
Idle Time (Rejected Material)
Idle Time (Pump Breakdowns)
LABOUR (CONCRETE PLACING TEAM)
Accidents
Pump Breakdown
Site Unprepared
Workers Skill Levels
Excessive Truck Mixers in System
Excessive Idle Time of Truck Mixers
TRUCK MIXER SURCHARGE

Figure 4

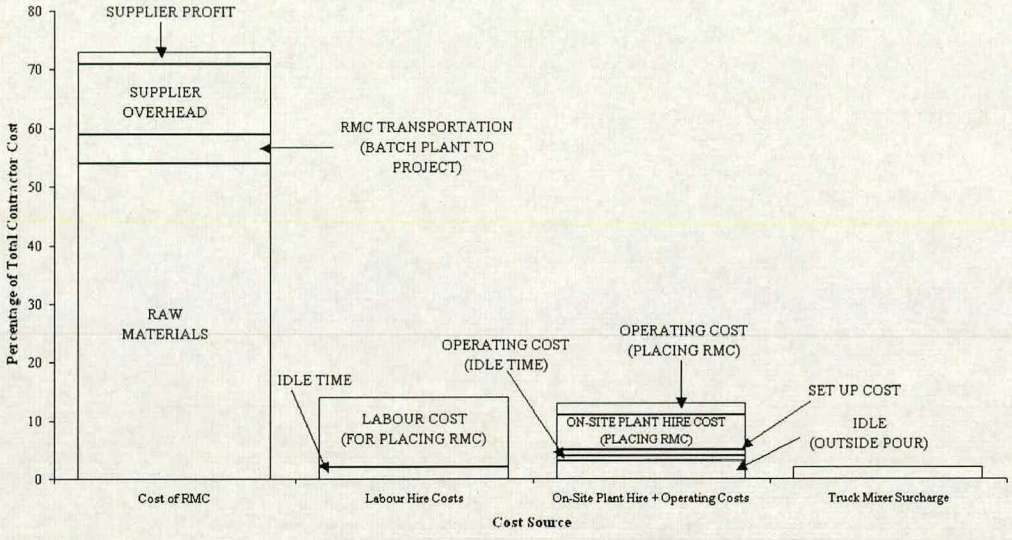

SUPPLIER PROFIT
SUPPLIER OVERHEAD
RMC TRANSPORTATION (BATCH PLANT TO PROJECT)
RAW MATERIALS
IDLE TIME
LABOUR COST (FOR PLACING RMC)
OPERATING COST (IDLE TIME)
OPERATING COST (PLACING RMC)
ON-SITE PLANT HIRE COST (PLACING RMC)
SET UP COST
IDLE (OUTSIDE POUR)
Percentage of Total Contractor Cost
80
70
60
50
40
30
20
10
0
Cost of RMC
Labour Hire Costs
On-Site Plant Hire + Operating Costs
Truck Mixer Surcharge
Cost Source

Figure 5

Data Collected
date,
structure location
type of pour;
weather (qualitative);
truck mixer ID;
time that batching started and finished;
travel time for truck mixer to site;
arrival time of a truck mixer at site;
discharge start time;
discharge completion time;
time washout started and finished;
time truck mixer left project site;
travel time for truck mixer to batch plant;
time of arrival at batch plant;
rejected loads;
time lost due to rejected loads;
time lost due to pump breakdowns;
wait period (if any) of truck mixers at batch plant, pump and washout;
time to position truck mixers at pump;
idle time of plant and labour due to a gap in the supply of RMC;
time truck mixers spent on project site;
pour duration;
number of labour force in each role (e.g. plant operator);
batch size and cumulative volume of RMC placed;
the target slump and actual slump of each batch.

Table 1

Operation Details	Observed Values
Volume of RMC (m^3)	42
Duration (hours)	3.72
No. Days for Operation (days)	1
Unit Costs	
Plant Hire (£/day)	345
Plant Operational Costs (£/hour)	20
Ganger (Foreman) Unit Cost (£/hour)	15
Concrete Placer Unit Cost (£/hour)	10
General Laborer Unit Cost (£/hour)	9.5
Plant Operator Unit Cost (£/hour)	14
RMC Material Unit Cost (£/m^3)	35
Resource Levels	
No. of Gangers (Foremen)	1
No. of Concrete Placers	5
No. of General Laborers	2
No. of Plant Operators	1
Truck Mixer Time at Site (mins)	150
Tolerance Threshold (mins)	=42*5 = 210

Table 2

Table 3

Figure and Table Captions

Figure 1: Observed pour volume histogram

Figure 2: Example of a recorded RMC operation

Figure 3: RMC supply and placing system

Figure 4: Ishikawa diagram of contractor costs in RMC system

Figure 5: Quantification of contractor costs in RMC system (small pours)

Table 1: Data variables collected for use in study

Table 2: Example operation: details, unit costs and resource levels

Table 3: Cost breakdown (as a % of total) comparison with RS Means (after RS Means, 1998) and Spon's (after Spon's, 2005)

A hybrid model to improve the estimation of concreting operations

L. Darren Graham[1] and Simon D. Smith

School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, UK

Stochastic construction operations, such as concreting, have been estimated effectively in the past using simulation. However, this effective estimation is only possible if the input (frequently probability distributions) to a simulation model is accurate. This accuracy is difficult to attain, and has prevented the widespread application of simulation in industry. An artificial intelligence technique, Case-Based Reasoning (CBR) is proposed as a method of improving the accuracy of input to a simulation model, allowing better estimates of concreting operations to be made. A simulation model, MatSim and a hybrid CBR-simluation model, CBRSim, have been developed (based on real construction data), validated, and compared to measure any improvements in output as a consequence of using a CBR-based input. CBRSim produced results that were more accurate and consistent than those of MatSim, indicating the potential advantage of using a CBR-based input in a simulation model of concreting operations. CBRSim is useful for the planning of concreting operations and may increase the industrial use of simulation as an effective estimation tool of stochastic construction operations.

Keywords: concreting operations, simulation modelling, case-based reasoning, construction planning.

INTRODUCTION

The concrete supply, delivery and placement process (concreting operations) is stochastic in nature and is often inefficiently managed. Concreting operations require analysis using non-deterministic techniques (Smith, 1998a) such as discrete-event simulation, which has been used to estimate the output of construction operations for over three decades (Martinez and Ioannou, 1999). Simulation has been proven as an effective tool for improving construction process planning (Halpin, 1993), but unfortunately the wide application of the technology in industry has been prevented by usage difficulties (Shi and Abourizk, 1997). One such difficulty is: in constructing and using a simulation model an understanding is required of the stochastic nature of the process concerned (Zhang et al., 2002). This nature is frequently represented by probability distributions and the importance of their effect on the accuracy of simulation output cannot be overemphasised (Abourizk et al. (1994), Banks and Carson (1984)). Therefore, to achieve an accurate simulation output, the probability

[1] Tel: 0131 650 5812, e-mail: Darren.Graham@ed.ac.uk

distributions must describe the stochastic nature of construction operations as closely as possible, and therein lies the difficulty: How does a planning engineer with little statistical knowledge determine the correct probability distributions and parameters that represent a particular construction operation? This paper presents an artificial intelligence technique, Case-Based Reasoning (CBR) as a possible solution to this problem. It is proposed that CBR could be used to: reduce the level of statistical knowledge required to use simulation; and to improve the input to, and consequently the output of, a simulation model.

Two models of the concreting system are presented and validated: a discrete-event simulation model, MatSim; and a hybrid CBR-simulation model, CBRSim. Both models: are based on probability distributions that have been derived from real construction project data; and, model the overall duration of a concrete pour. Before outlining the models, this paper will look briefly at the process of concrete supply, delivery and placement.

THE CONCRETING SYSTEM

In previous research, concreting operations have been separated into two processes: batching and delivery (for example, Sawhney et al., 1999); and pumping and placing (Dunlop and Smith (2000 and 2002)). In this study, the concrete supply, delivery and placing process is considered in its entirety. In an ideal situation, a truck mixer is filled with concrete at the batching plant and travels to the construction site. On arrival, the truck mixer moves into position at the pump and discharges its load. Finally, the empty truck mixer is cleaned in the washout area and proceeds back to the batching plant, completing the cycle. In reality, the events that occur within the system (e.g. truck arrival times and pump start times) take place at irregular intervals (Smith, 1998b). Queuing of trucks at the batching plant, pump and washout areas can be expected, as it is often unlikely that a truck will arrive simultaneous to the previous truck departing. If trucks arrive late, there will be a lengthening of the process and under-utilisation of the pump- resulting in an inefficiently managed pour of the concrete. The stochastic nature of the concreting system can be recreated by using probability distributions to represent the following times of interest: travel time from batching plant to construction site (travel to site time); positioning time at pump (position time); discharge time (pump time); time to clean the truck mixer (washout time); travel time from site to batching plant (travel to batcher time); and load time (batch time). The probability distributions can produce times that are used to recreate a typical cycle of the process. Through the simulation of multiple cycles, the attributes of a particular operation's set-up can be provided, such as pour duration (from which productivity can be derived) or pump utilisation rates. To ensure that the model is realistic and accurate, it must be based on data collected from real construction sites.

DATA COLLECTION

The data used to build both MatSim and CBRSim were collected during a time study of four construction projects, which are detailed in Table 1. In total, 225 concrete

pours were observed and the variables of interest were recorded. These variables were: pour date and type, weather, target and actual slump, batch time, arrival on site time, position time, start and completion of pumping times, washout time, batch quantity, total pour volume, truck wait times (at batcher and pump), batcher and pump idle times, and the number of trucks involved in the pour. From the observed data, the times of interest mentioned previously were extracted.

The data that were used in validation were collected from a construction project independent from those used to build the models (See Table 1). In total, 63 pours were observed and from this data set, 20 pours were selected at random for use in the validation process.

Project	Use in Study	Year of Completion	Type of Project	Location	No. of Observed Pours
1	Build Model	2000	Wastewater Treatment Plant Construction	Aberdeen, UK	152
2	Build Model	2000	Wastewater Treatment Plant Construction	Peterhead, UK	10
3	Build Model	2000	Wastewater Treatment Plant Construction	Fraserburgh, UK	18
4	Build Model	2000	Wastewater Treatment Plant Construction	Dundee, UK	45
5	Model Validation	1994	Motorway Viaduct Strengthening and Widening	Cheshire, UK	63

Table 1: Details of studied projects

DISCRETE-EVENT SIMULATION MODEL, MATSIM

MatSim is a computer model written in the Matlab programming language and is based upon the discrete-event simulation methodologies described fully in Smith (1998a and 1999) and Law and Kelton (2000). To summarise the steps involved in the developing the model were:

- Fitting probability distributions. The theoretical probability distributions that best represent the previously mentioned times of interest were identified using a commercial package, BestFit. The fit of the distributions to the data was assessed using the K-S test and visual assessment. Generally, the data sets were best represented by the lognormal, gamma and beta distributions and only they will be considered further in MatSim.
- Generation of random variates. A random variate is a value (in this case a time of interest, e.g. position time) generated at random from the chosen probability distribution. Assuming a good fit between the data and the probability distribution, the random variate is a true representation of an actual value (Dunlop and Smith, 2000). A full discussion on the methods of generating random variates can be found in Law and Kelton (2000).
- Use random variates to synthesise 'events'. An event is something that changes the state of the concreting system. It could be an arrival or departure at the batching plant, pump or washout area.

- Model operations. A real operation is a series of events, the timing of which determine its attributes (Dunlop and Smith, 2000) (for example, pour duration). These attributes can be determined for a simulated operation and to reduce variance, many replications are performed.

HYBRID CBR-SIMULATION MODEL, CBRSIM

CBR is the process of reasoning and learning by storing cases- records of specific prior reasoning episodes- and retrieving and adapting them to aid new problem solving or interpretation in similar situations (Kolodner, 1993). Aamodt and Plaza (1994) described the CBR process as cyclical, comprising of the following:

- Retrieval of the most similar case(s) from the case base
- Reuse of the case(s) to attempt to solve the problem
- Revision of the proposed solution, if necessary, to solve the problem
- Retention of the modified solution as a new case

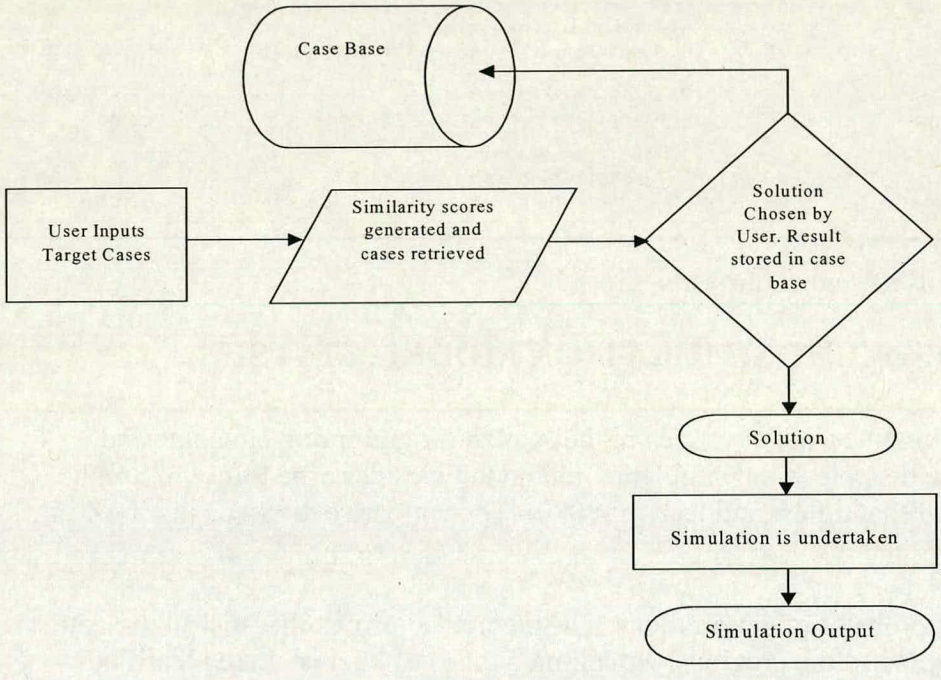


Figure 1: Flowchart of CBRSim model

The construction domain is a growing area for applications of CBR and it has been used successfully in solving prediction and estimating problems (Morcous et al., 2002). This study intends to utilise CBR to predict the probability distributions and parameters that best represent a particular concreting operation. The predicted distributions will subsequently be used in a simulation model, forming the hybrid CBRSim model. CBRSim is based upon the cyclical process described by Aamodt and Plaza and its methodology is shown in Figure 1. In developing CBRSim the stages were:

(a) Case base design
(b) Case retrieval

(c) Simulation using CBRSim

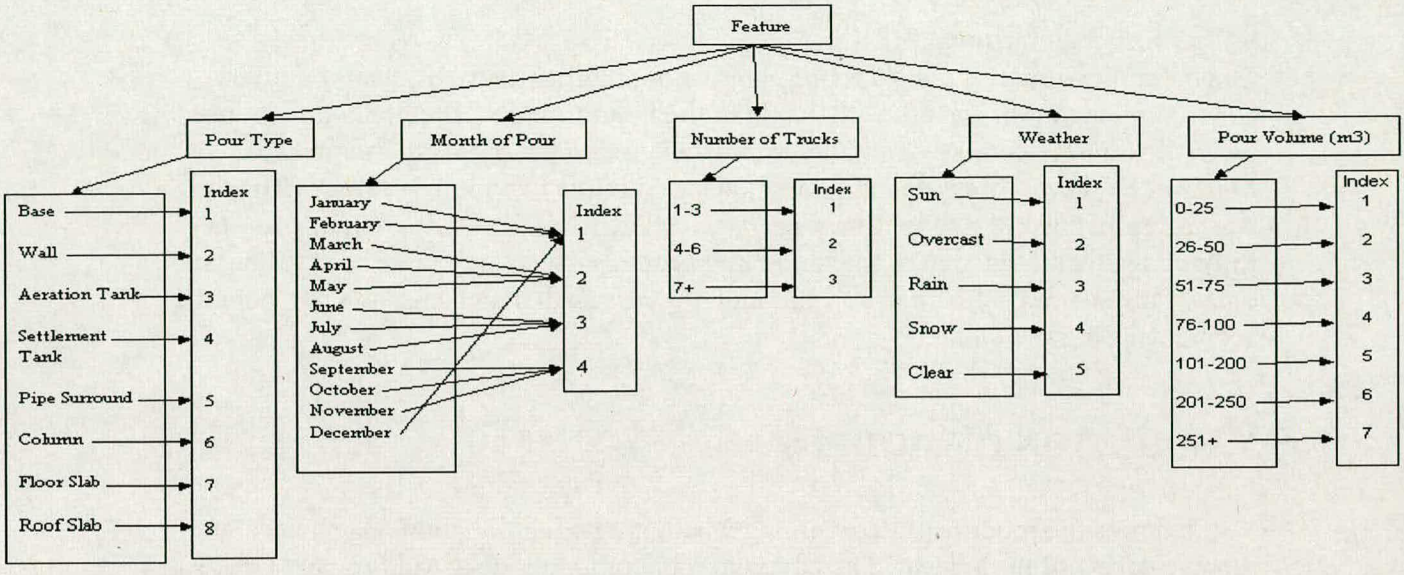


Figure 2: Feature indexing system in the CBRSim model

(a) Case Base Design

The case base is the storage area for prior cases. Cases must be parameterised, typically by being divided into a series of distinct features by means of which their similarity to other cases can be judged. In CBRSim, the case base initially stored 225 cases, with all the cases sharing the same record structure of 5 input and 5 output features. The input features are: the number of trucks, pour volume, pour type, month of pour and weather. Each feature is assigned an index to aid the retrieval process (See Figure 2). The output features are the probability distributions and parameters that best represent each concreting operation, identified using BestFit. Seven distributions provided suitable representations of the concreting process, they were: uniform, normal, lognormal, exponential, weibull, gamma and beta. The next stage of the CBR process is case retrieval.

(b) Case Retrieval

A CBR model derives its power from its ability to retrieve relevant cases quickly and accurately from its case base (Arditi and Tokdemir, 1999). To perform the retrieval process, CBRSim measures the similarity between a user specified target case and the indices of each case stored in the case base. This similarity measurement is in percentage form and is based upon Euclidean distance measurements, as detailed in Hand et al. (2001). If the similarity between the target case and a stored case is greater than 75%, the stored case is retrieved from the case base and is placed in a temporary storage area. Once all the relevant stored cases have been placed in the temporary storage area, the user is permitted to view them. The user is then required to select a case based only on the similarity measurements, removing the need for a user to have statistical knowledge of the process. If the target case features are not identical to the features of the selected case, CBRSim retains the target case features and the probability distribution information as a new case. Thus providing the learning ability of the model and completion of the CBR cycle. Finally, the case base should be updated over time with the details of new construction projects, improving the knowledge of the model.

(c) Simulation using CBRSim

The determination of the most representative probability distributions for a given concreting operation ends the CBR side of the hybrid model. The methodology used by CBRSim to undertake simulation of the concreting process is the same as in MatSim. However, the computing methods used in the two models differ. CBRSim was developed using a spreadsheet package, Microsoft Excel, allowing the user to fully understand each step in the modelling process. This study must now validate both MatSim and CBRSim. The validation process shall also be used to compare the accuracy of the two models.

VALIDATION OF MODELS

Validation is the process of determining whether a simulation model is an accurate representation of the system. The concreting operation models, MatSim and CBRSim are intended to represent the existing system, not replace it, in order to allow estimations of the system output to be produced. Therefore, the model output (overall duration of a concrete pour) and data from the existing system can be compared to ascertain if the model is valid. In this study, a correlated inspection approach was used to indicate the validity of both models. Correlated inspection at its most extreme requires that a model be driven by exactly the same observations as the real system, but it is unclear if this indicates the predictive powers of a model (Smith et al., 1995). Therefore, the model will be driven with random variates generated from probability distributions.

20 concrete pours (validation observations) were randomly selected from the validation data set discussed earlier. MatSim was investigated when random variates were generated from the lognormal, gamma and beta probability distributions. The parameters of these distributions were tailored to each individual pour, using knowledge of the stochastic nature of the process. The performance parameter of interest in this study is the pour duration and the results are shown in Figure 3. To overcome the variance in simulation output the model's results were replicated 100 times, and an average result was taken.

Validation Observation	Overall Pour Duration (mins)	Percentage difference from observed value			
		CBRSim	MatSim Lognormal	MatSim Gamma	MatSim Beta
1	230	-1.5	-16.4	-6.3	-2.5
2	316	3.0	10.8	21.3	22.4
3	244	0.7	-9.3	0.4	5.7
4	351	-0.8	-22.1	-19.9	-17.8
5	396	-0.9	-2.5	8.3	15.4
6	318	1.0	-23.3	-18.9	-22.1
7	272	-3.5	-10.0	-7.6	-5.7
8	303	-2.4	-18.5	-15.8	-13.5
9	273	-1.7	-23.4	-21.3	-17.0
10	331	0.6	-19.3	-14.4	-12.6
11	409	-1.7	-18.0	-15.3	-12.6
12	288	-0.3	-8.8	-10.3	-6.3
13	235	2.8	-12.5	-8.6	-5.0
14	223	2.0	-7.8	-4.9	1.5
15	294	-1.0	-9.9	-4.0	-0.9
16	199	2.4	2.7	1.3	6.7
17	280	0.0	-7.0	-8.9	-3.9
18	225	-1.0	-22.5	-18.9	-14.8
19	255	1.2	19.3	-12.7	-11.9
20	477	0.2	16.5	11.9	14.2

Table 2: Correlated inspection results for pour duration

Comparison of Results

It was the aim of this study to determine if a CBR derived input could improve the output of a simulation model. This improvement should be measured in terms of accuracy and reliability of the model. An examination of the results presented in Figure 3 shows that CBRSim has the ability to provide better estimates, generally, than MatSim. CBRSim achieved pour duration estimates to within a range of +/- 3% of the observed values, compared with a range of +/- 20% produced by MatSim. In terms of reliability, CBRSim estimated to within 10% of the observed value in 20 out of 20 validation observations (Table 2). In contrast, MatSim had a poor reliability record for pour duration estimating less than 50% of validation observations within 10% of the observed values (Table 2). In summary, both models provide reasonable estimates of the pour duration and can be affirmed as valid.

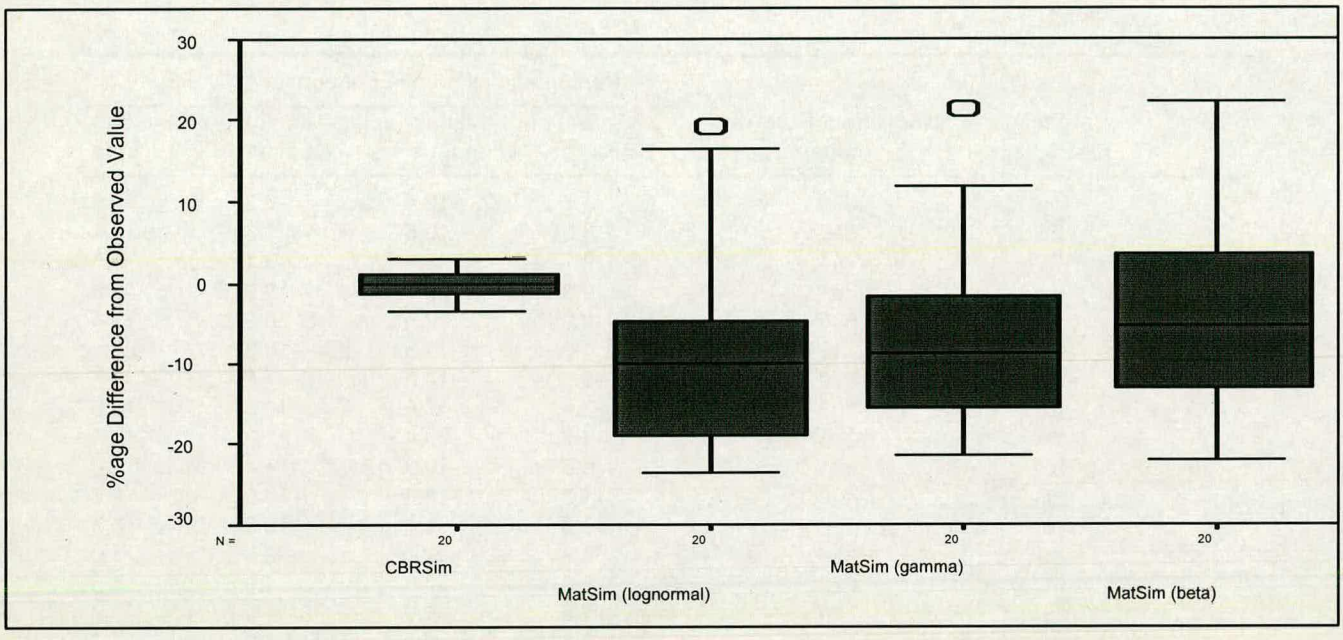


Figure 3: Box plots of percentage difference between model and observed values

CONCLUSIONS

In the past, simulation has been used to estimate the output of construction operations and has proven to be an effective tool for planning random construction processes, such as concreting. However, simulation is a difficult tool to use, as the output of a simulation model will be inaccurate if the input is unrepresentative of the environment. When probability distributions are the input to simulation models, their characteristics can be estimated using CBR. This removes the need for the user to have statistical knowledge of these distributions. A simulation model, MatSim and a hybrid CBR-simulation model, CBRSim were developed based on real construction project data. They were validated using data from a construction project independent to those on which the models were built. The pour duration estimates from both models were compared to determine the effect of using CBR in this study. It was found that CBRSim was able to estimate pour durations to within +/- 3% of the observed pours, compared with +/- 20% from MatSim. In terms of reliability, CBRSim estimated to within +/- 10% of the observed values in 100% of the validation observations. MatSim was unable to match this reliability record. It may be concluded that a case-based reasoning input simplifies the use of simulation, whilst improving its accuracy-making CBRSim a useful tool for the planning of concreting operations. The simplification of simulation modelling may increase the industrial application of simulation as an effective estimation tool of stochastic construction operations.

REFERENCES

Aamodt, A., and Plaza, E. (1994) Case-based reasoning: foundational issues, methodological variations, and systems approaches, *AI Communications*, 7(i), 39-59.

Abourizk, S.M., Halpin, D.W., and Wilson, J.R. (1994) Fitting beta distributions based on sample data, *J. Constr. Engrg. And Mgmt., ASCE*, 120(2), 288-305.

Arditi, D., and Tokdemir, O.B. (1999) Comparison of case-based reasoning and artificial neural networks, *J. Comp. In Civil. Engrg., ASCE*, 13(3), 162-169.

Banks, J., and Carson, J.S. (1984) *Discrete-event system simulation*, Prentice-Hall Inc., New Jersey, USA.

Dunlop, P., and Smith, S. (2000) Stochastic modelling of concrete operations, *In:* Akintoye, A (Ed.), *16th Annual ARCOM Conference*, 6-8 September 2000, Glasgow Caledonian University. Association of Researchers in Construction Management, Vol. 2, 483-492.

Dunlop, P., and Smith, S. (2002) Simulation analysis of the UK concrete delivery and placement process: a tool for planners, *In:* Greenwood, D (Ed.), *18th Annual ARCOM Conference*, 2-4 September 2002, Northumbria University. Association of Researchers in Construction Management, Vol. 2, 781-790.

Halpin, D. W. (1993) Process-based research to meet the international challenge, *J. Constr. Engrg. And Mgmt., ASCE*, 119(3), 415-425.

Hand, D., Mannila, H., and Smyth, P. (2001) *Principles of data mining,* MIT Press, Cambridge, Mass., USA.

Kolodner, J. (1993) *Case-based reasoning,* Morgan Kaufmann, San Mateo, Calif., USA.

Law, A. M., and Kelton, W. D. (2000) *Simulation modelling and analysis*, 3rd edition, McGraw-Hill Inc, New York, USA.

Martinez, J. C., and Ioannou, P. G. (1999) General-purpose system for effective construction simulation, *J. Constr. Engrg. And Mgmt., ASCE*, 125(4), 265-276.

Morcous, H., Rivard, A. M., and Hanna, A. M. (2002) Case-based reasoning system for modelling infrastructure deterioration, *J. Comp. in Civil. Engrg., ASCE*, 16(2), 104-114.

Sawhney, A., Abudayyeh, O., and Chaitavatputtiporn, T. (1999) Modeling and analysis of concrete production plant using petri nets, *J. Comp. in Civil. Engrg., ASCE*, 13(3), 178-186.

Shi, J., and Abourizk, S.M. (1997) Resource-based modeling for construction simulation, *J. Constr. Engrg. and Mgmt., ASCE*, 123(1), 26-33.

Smith, S.D., Osborne, J.R., and Forde, M.C. (1995) Analysis of earth-moving systems using discrete-event simulation, *J. Constr. Engrg. and Mgmt., ASCE*, 121(4), 388-396.

Smith, S. D. (1998a) Concrete placing analysis using discrete-event simulation, *Proc. Institution of Civ. Engineers, Structures and Buildings*, 128 (4), 351-358.

Smith, S. D. (1998b) Stochastic analysis of cyclic construction processes, *In:* Hughes, W. (Ed.), *14th Annual ARCOM Conference*, 9-11 September 1998, University of Reading. Association of Researchers in Construction Management, Vol. 1, 127-136.

Smith, S.D. (1999) Modelling and experimentation of the concrete supply and delivery process, *Civ. Eng. And Env. Syst.*, 16 (2), 93-114.

Zhang, H., Shi, J.J., and Tam, C.M. (2002) Iconic animation for activity-based construction simulation, *J. Comp. in Civil. Engrg., ASCE*, 16(3), 157-164.

A METHOD FOR EFFECTIVELY IMPLEMENTING CONSTRUCTION PROCESS PRODUCTIVITY ESTIMATION MODELS

L. Darren Graham[1] and Simon D. Smith

School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, UK

The international research community have proposed many accurate models for estimating the productivity of construction processes. Few of these models permit practical implementation at the early stages of a construction project, and therefore are of limited use to construction practitioners. For example, contractors require accurate estimates of productivity in the early stages of a project, to avoid being lumbered with an inadequate quantity of plant at the construction stage.

This paper considers the earthmoving process, and examines the effect of inadequate resources on productivity. The causes of under-resourced operations are discussed and a dynamic modelling framework is proposed as a method to prevent this phenomenon. Dynamic modelling can be applied to any modelling method and constitutes: separating the operation into planning stages; choosing the variables which are known and significant at each stage; building a model based on those variables. The framework is applied and validated for a linear regression model of earthmoving productivity, providing acceptable results. The accuracy of the results from the model could be improved by the adoption of the principles of concurrent engineering. Finally, the implementation of dynamic modelling in simulation, case-based reasoning and neural networks is discussed.

Keywords: estimating, planning, modelling, productivity

INTRODUCTION

There have been numerous modelling methods, such as simulation, neural networks, regression analysis and case-based reasoning, proposed to estimate the productivity of construction processes. The productivity is normally a measure of the amount of work completed per hour, and can be used to estimate the amount of time and financial cost required to complete an individual operation of a construction process. The time and cost estimates for each operation, of every construction process are required to make a plan of a complete project, and therefore these estimates must be accurate. In the past there have been many simulation productivity models, such as: CYCLONE (Halpin, 1977), UMCYCLONE (Ioannou, 1989), STROBOSCOPE (Martinez and Ioannou, 1994) and TruckSim (Smith, 1995; Smith et. al, 1995). More recently the tendency has been to use the artificial intelligence modelling techniques neural networks (Chao and Skibniewski, 1994; Portas and Abourizk, 1997; Sonmez and Rowings, 1998) and case-based reasoning

[1] Tel: 0131 650 7207, e-mail: darren.graham@ed.ac.uk

(Graham and Smith, 2003; Graham et. al, 2004; Graham and Smith, 2004) to estimate construction productivity.

Although, all of the above models have been proven capable of providing accurate estimates of productivity, there have been many associated usage difficulties, preventing the widespread application of the models in the construction industry. One of the main difficulties lies in the need to provide a supplier (of plant and materials) with an accurate order of the requirements for a particular operation (or operations) far in advance of the day of construction. This holds for any large UK Contractor, whether the supplier is internal or external to an organisation. The productivity models proposed so far have not taken into account the need for early, accurate estimates and hence are less useful to industry than they could be.

A consequence of placing a late order with a supplier, especially of plant, is to potentially leave an operation under resourced. To highlight the effect that a lack of resource has on operational performance, a study is presented to quantify the changes in performance encountered in an operation, resulting from a reduction in resources. The consequences of recruiting too much plant (over resourced) are also discussed. Next, a method of making productivity models more useful is presented, based on the philosophy that a model should be dynamic. That is, the variables which are required to be input by the user (planning engineer) should be realistically known at pre-defined stages of a construction project; the model variables change (normally increase in number) as the time to construction draws nearer and a planning engineer has more knowledge of an operation. In this paper, the dynamic modelling framework has been applied to the productivity estimation of earthmoving operations using linear regression analysis, at two project stages.

The structure of this paper is as follows:

- Summary of the collection of data
- The effect of poor plant allocation on construction operations
- The dynamic modelling framework
- An example application of dynamic modelling: linear regression analysis of earthmoving productivity
- Re-planning the validation operations
- Implementing dynamic modelling using other modelling methods
- Conclusions and future work

GATHERING OF REAL CONSTRUCTION DATA

The data used in this study was collected during a time study of four road-building projects, undertaken by the same contractor in UK. These projects varied widely in terms of: the quantity of soil that was required to be moved, from the relatively small A52 Ashbourne relief road (120,000 m^3) to the M1/A1 link road (3,000,000 m^3); the location of the projects (South, Midlands and North of England); the soil types encountered; and numerous additional factors. In total, 141 earthmoving operations were recorded, 90 of which were deemed suitable for use in this study; the 51 others were discounted due to

missing data. Two-thirds of the 90 operations were selected for use in model development, the remaining third to be used in model validation. Table 1 presents a list of the recorded variables and their corresponding determinability at: four to six weeks prior to construction (planning stage one); and, one week to one day before construction (planning stage two). If a variable could be determined with a high degree of confidence (high in the table) it was considered for use in modelling (see table 1).

THE EFFECT OF POOR PLANT ALLOCATION ON CONSTRUCTION OPERATIONS

It is important to highlight the effect that the under-provisioning (poor allocation) of plant has on the performance of construction operations. The aim of this study is to quantify this effect, and will do so by:

- using a linear regression model to recreate the thirty validation operations which have taken place in the past, reducing the number of haulers used in the operations by increments of one, until the final recreated operation uses three haulers less than the original. Recording the performance of each new operational set-up.
- measuring the difference (percentage) between the original performance and each modelled operational performance to gain some perspective on the effect of reducing the quantity of plant.

Table 1: Recorded variables and variables considered for modelling

Variable	Determinability		Model Variables	
	Planning Stage One	Planning Stage Two	Planning Stage One	Planning Stage Two
Haul road soil parameters	Medium	High	Haul road gradient	Haul road soil parameters
Haul road gradient	High	High	Plant specifications	Haul road gradient
Haul road rolling resistance	Low	Medium	Loader bucket volume	Haul road length
Haul road length	Medium	High	Number of haulers used	Plant specifications
Plant specifications	High	High	Total volume to be moved	Load cycle times
Plant operator ability	Low	Medium	Bucket passes per load	Hauler travel times
Load area characteristics	Medium	Medium	Project number *	Loader bucket volume
Type and load of obstructions	Low	Medium		Bucket passes per load
Weather	Low	Low - Medium		Number of haulers used
Material susceptability to weather	Low	Low - Medium		Total volume to be moved
Load cycle times	Medium	High		Month of excavation
Hauler manoeuvre times	Medium	Medium		Project number *
Hauler travel times	Medium	High		
Load dump times	Medium	Medium		
Loader bucket volume	High	High		
Bucket passes per load	High	High		
Number of haulers used	High	High		
Total volume to be moved	High	High		
Month of excavation	High	High		

* included to allow an examination of the effect that an individual project has on all the collected data. If the effect is significant, the model will not be useful for providing estimates of new projects.

The results of the poor plant allocation study are shown in table 2; they are an average over the entire validation data set. Clearly, a reduction in the number of haulers causes a steady decline, of 16.21% per truck of operational productivity and a non-linearly increasing operational cost. Therefore, it is very important to ensure that an adequate

quantity of plant is available for a specific operation, and the dynamic modelling method is one way of providing this.

Table 2: Average percentage difference from the observed, for the whole validation set

	Number of Haulers Removed		
Performance Measure	One	Two	Three
Productivity	-16.21	-32.42	-48.63
Cost	7.53	12.01	23.70

THE DYNAMIC MODELLING FRAMEWORK

The study in the previous section highlighted the detrimental effect of having too few trucks in an earthmoving operation. What causes an under-resourced operation? There are a number of answers to this question: an inaccurate estimate of the probable rate of work (productivity) is made; an order for plant requirements is submitted to the supplier too late to allow fulfilment. What can be done to reduce the occurrence of an under-resourced operation? To diminish these issues, two methods could be pursued:

- Concurrent engineering principles should be adopted by construction firms. Briefly, this means bringing together all parties: designer, architect, client, construction planner and contractor to learn as much about the project at the earliest possible stage. This shall not be dealt with here.
- The dynamic modelling framework should be applied to productivity estimation models. This shall now be discussed in full.

The aim of dynamic modelling is to provide an accurate estimate of the productivity of a construction operation at the earliest possible stage in the planning of a construction project. This aim is achieved through using information about the operation that is realistically known at certain stages of the planning process. Currently, the dynamic model is split into two stages: planning stage one, four to six weeks prior to construction, with the aim of allowing a planning engineer to calculate the resource requirements; planning stage two, one week to one day prior to construction, to allow a confirmation of the order or make a slight adjustment to the order based on new information.

The framework of the dynamic modelling philosophy is:

- Make a list of all available variables and determine how if they are known at either planning stage one or two. If a variable is known it is added to a second list.
- Select the significant variables from the second list for each planning stage; principle component analysis (PCA) was used to choose these variables in the study.
- Make two models (one for each planning stage) based on the significant variables.
- Validate the model to provide confidence that it provides accurate estimates.

- At planning stage one; supply the known details of the operation to the model, and experiment with the number of resources to produce estimates of productivity and operational cost. Compare the productivity and cost estimates and note the plant resource which provides the best value for money, e.g. a compromise between speed and cost. Increase the ideal numbers of plant by one to provide an initial conservative estimate. Communicate these requirements to the plant supplier(s).
- At the second planning stage: supply the required information to the model, to produce an adjusted estimate of productivity and cost; experiment with the plant requirements to find the compromise between speed and cost. At this stage the ideal number of plant should be chosen and communicated to the plant supplier(s). It is unlikely that the plant requirements will have changed significantly, therefore the suppliers should be able to meet the contractor's needs.

AN EXAMPLE APPLICATION OF DYNAMIC MODELLING: LINEAR REGRESSION ANALYSIS OF EARTHMOVING PRODUCTIVITY

This example demonstrates the application of linear regression to modelling earthmoving productivity through following the dynamic modelling framework outlined above. The data used in this experiment was introduced earlier in this paper. To re-cap, 90 operations were observed, 60 of which shall be used in model development, the remainder for validation.

Select the highly determinable earthmoving process variables
The highly determinable variables of the earthmoving process have been selected and are shown in table 1. Note there are 6 and 11 such variables, at planning stages one and two, respectively. In addition a project number index has been included to determine whether the models only hold true for specific projects- if this were the case the models' predictive usefulness would be limited; this results in, 7 and 12 process variables at planning stages one and two, respectively.

Select the significant variables of the earthmoving process
PCA was used to select the significant variables of the process. The numbers of variables were reduced from: 7 to 5 at planning stage one; 12 to 9 at planning stage two. For both planning stages, the project index was insignificant, thus the models will be useful for providing predictions of future operations.

Develop a model for each planning stage
A linear regression model was developed for both planning stages one and two, using the remaining data. The regression method used was the backward elimination stepwise, meaning that a variable which was found to be less significant in the linear fit than a threshold value, would be removed. The regression is then performed again on the reduced data set, and this process continues until the remaining variables are above the set threshold of significance. The planning stage one variables were reduced from 5 to 3, and stage two variables reduced from 9 to 6.
The regression equations for the productivity are:

Productivity = 18(No. Haulers) + 10.8(Hauler Type) + 0.3(Volume) + 44.5 Stage One

Productivity = -14.5(Month) + 9.6(Loader Type) +25.5(No. Haulers) – 40(Load Cycle Time) + 27.2(Buckets per Load) – 7.2(Travel Time) + 147.7 Stage Two

How well does the model fit the data? A coefficient of variation in the model of 1 signifies a perfect fit. However, a value of 0.7 is most likely to be an acceptable fit to a complex system like the earthmoving process. The R^2 values for the planning stage one and two models were 0.755 and 0.88, respectively. Thus, these regression fits can be deemed to be tentatively acceptable. There is a need for more knowledge of the predictive capabilities of the models, and this shall be provided through model validation (see next indent).

Model validation

The details of 30 observed operations, which had not been used in the model development, were used in the model validation. Validation consisted of entering the information of each operation into both regression models and recording the estimated productivity values. These values were compared with the observed productivities and it was found that on average: the planning stage one model overestimated productivity by 5.8%, with a standard deviation in results of 27.7%; the planning stage two model overestimated productivity by 3.13%, with a standard deviation in the results of 14.3%. Thus, due to more information being known about the process at the later stage, the model of stage two is more accurate. A plot of the models' estimates against the observed productivities for the validation set is shown in figure 1. Clearly, the models are roughly following the trend of the observed data. In conclusion, both models are suitable accurate for the purposes of this study.

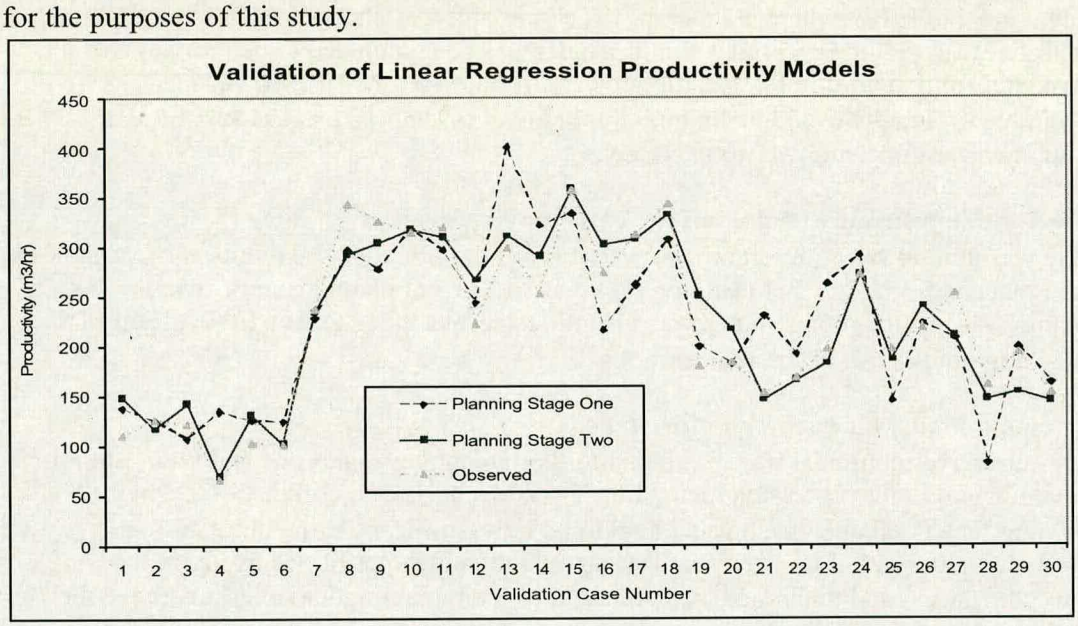


Figure 1: The results of model validation

Using the models to order plant

The known details of a future operation can be supplied to the validated models, in order to receive estimates of productivity, duration and cost. The number of plant required (number of haulers) can be experimented with (between 2 and 6 haulers in this study) and a compromise between duration and cost of an operation can be made to find the ideal planning solution. In this example the details of the operation, known at planning stage one were: Type of available hauler = Caterpillar D400D; Excavation volume = 400 m^3. At planning stage two, the known details of the operation were: Month of excavation = May; Type of available loader = Caterpillar 245 backhoe excavator; Load cycle time = 2 minutes; Buckets per load = 7; Travel time = 6 minutes. The number of haulers was experimented with, and the resulting productivity and cost estimates for each experiment, at planning stages one and two, are shown in tables 3 and 4, respectively. Note that in this example the trend of the costs observed in tables 3 and 4 are the reverse of those in table 1, the consequence being that this operation provides more value for money when only two haulers are used, and it is likely that poor forward planning would have resulted in this operation having been over-resourced, i.e. 3 haulers used instead of 2.

Table 3: Experimental results at the early planning stage (one)

	Number of Haulers				
	2	3	4	5	6
Productivity (m3/hr)	210.2	228.1	246.1	264.1	282
Cost (£)	216.65	271.98	319.24	360.06	395.67
Cost/ Volume (£/m3)	0.537	0.675	0.792	0.893	0.982

To make the decision on which operational set-up provides the best value for money a ratio of the cost divided by volume is calculated. As volume is fixed the ratio provides a basis for directly comparing the results of the two planning stages and the value for money of each resource quantity. In the cost/volume ratio, a smaller value signifies a better compromise between cost and time. Clearly, at the first planning stage, having two haulers provides the best value for money. Therefore, following the dynamic modelling framework, an order should be placed at this time, of four to six weeks to construction, with the plant supplier for three haulers (ideal plus one). Examining the experimental results at the second planning stage reveals that the situation is unchanged in terms of the ideal quantity of haulers. However, it is noteworthy that the value for money for two haulers has been reduced (cost/volume ratio increased) in the second stage, while the remaining hauler quantities provide increasing value (cost/volume ratio decreased).

Table 4: Experimental results at planning stage two

	Number of Haulers				
	2	3	4	5	6
Productivity (m3/hr)	209	234	260	285	310
Cost (£)	217.97	264.75	302.36	333.25	359.08
Cost/ Volume (£/m3)	0.541	0.657	0.75	0.83	0.89

RE-PLANNING OF VALIDATION OPERATIONS

Each of the 30 validation operations discussed in the previous section, were re-planned using the methods outlined above, to produce new estimates of cost and productivity. By comparing the cost estimates with those actually observed in the operations, it is possible to calculate the average improvement/deterioration in the cost (per cubic metre) to complete an operation, over the whole validation set, from using the dynamic modelling framework to plan the operations. It was found that an average cost saving of £0.21 per cubic metre was possible.

Assuming that the validation set of operations is an appropriate survey of the four studied projects, the cost saving per cubic metre of earth moved can be used to calculate the possible cost savings in all four projects. The total amount of earth moved in the four studied projects equalled 5.82 million cubic metres, and the actual cost of this work was calculated as £5.98 million. Through re-planning using the dynamic modelling framework the total cost could have been £4.75 million, offering a reduction in cost of 20.5% from the actual. It is noteworthy that this is only a theoretical saving to the direct cost of earthmoving- indirect costs were not considered. Clearly the saving indicates the potential impact that the dynamic modelling framework could have on a construction industry which needs to reduce waste and to improve its performance.

IMPLEMENTING DYNAMIC MODELLING USING OTHER MODELLING METHODS

The following section is a brief summary of the actions required to implement the dynamic modelling framework in other commonly used construction process modelling methods.

Case-Based Reasoning (CBR)

CBR consists of formulating a data (case) base of past examples of a problem and using several different methods to identify appropriate past examples as possible solutions to new problems. Thus, the implementation of dynamic modelling can be achieved in two ways: one, as suggested in the dynamic modelling framework, point 3- have a separate case base for each planning stage, incorporating only the significant variables of the respective stages; two, vary slightly from the framework, point 3 by building one case base holding the significant variables of both planning stages, but only consider, at any one time, the variables relevant to a particular stage of the project. The remainder of the framework can be followed using CBR.

Artificial Neural Networks (ANN)

ANNs form approximations of the trends contained within data sets and are therefore useful in modelling many non-linear or near-linear problems, such as predicting the productivity of the earthmoving process. ANNs work in a similar way to linear regression analysis and the dynamic modelling framework can be followed directly by: separating the data into two groups, planning stages one and two; splitting the data within each of these groups into two further groups- training and testing; training two models to mimic

planning stages one and two; testing the models predictive capabilities (validation) and then using them to plan future construction operations.

Discrete-Event Simulation (D-E Sim)

Discrete-event simulation models are often driven by probability distributions. These distributions represent the variability contained within the process, which is being modelled. In the earthmoving process, the variability can be described by representing the stochastic elements in the process, such as the time of travelling from the loader to the dump area (travel time), using probability distributions. Probability distributions that accurately represent the stochastic elements can then be used to build up a picture of the productivity of the process, over an entire operation. The probability distributions which represent the stochastic elements will change with different operational conditions, i.e. the probability distributions representing the travel time at planning stage one, are unlikely to remain the same at planning stage two. At the moment, to understand the nature of this change would require an expert. However, Graham and Smith (2003; 2004) have produced a computer model which removes the need to have expert knowledge of the system to define its representative probability distributions, and the dynamic modelling framework shall be applied to this model in the future.

CONCLUSIONS

A by-product of poor forward planning is that a construction project can be under-resourced. This is due to a plant/materials supplier not having a sufficient stock, at late notice, to meet the demands of a construction project. This study highlighted the effect of having under-resourced operations on the earthmoving process performance. It was found that for every hauler unit an operation was under resourced by, the productivity was reduced on average by 16%, with a concurrent non-linear increase in cost. Thus, the provision of too little resources should be avoided. Intuitively, providing too many resources should also be avoided, to prevent the incursion of additional costs- an extra hauler makes a large difference to the cost of an operation, and this can propagate for a whole project.

In an attempt to improve the planning of projects, and to prevent the issues raised above, this paper proposed that construction productivity models are developed following the dynamic modelling framework. This framework consists of using all the significant knowledge possessed by a planning engineer at specific project stages, to allow accurate estimates to be made of the resources required in an operation, well in advance of construction.

The dynamic modelling framework was applied to the linear regression analysis of the earthmoving process, providing models which accurately recreated a validation set of operations. Experimentation was undertaken on how the productivity and cost of the operation changed with a change in the number of haulers. The results of this experimentation allowed the validation operations to be theoretically re-planned, with an ideal number of haulers selected for each situation. This new estimate of performance, in terms of cost and duration was compared with the observed values. This comparison was

used to determine if any theoretical improvement could have been made, by applying the dynamic modelling framework to the planning of operations in the first place. It was found that a theoretical 20% saving on cost could have been made through applying the dynamic modelling framework.

REFERENCES

Chao, L-C., and Skibniewski, M. J. (1994) Estimating construction productivity: a neural network based approach, *J. Comp. In Civil. Engrg., ASCE*, 8(2), 234-251.

Graham, L. D., and Smith, S. D. (2003) A hybrid model to improve the estimation of concreting operations, *In:* Greenwood, D (Ed.), *19th Annual ARCOM Conference*, 3-5 September 2003, University of Brighton. Association of Researchers in Construction Management, Vol. 2, 505-512.

Graham, L. D., and Smith, S. D. (2004) Estimating the productivity of cyclic construction operations using case-based reasoning, *Advanced Engineering Informatics*, In Print.

Graham, L. D., Smith, S. D., and Crapper, M. (2004) Improving concrete placement simulation with a case-based reasoning input, *Civil Engineering and Environmental Systems*, Vol. 21(2), 137-150.

Halpin, D. W. (1977) CYCLONE: method of modeling job site processes, *J. Constr. Div., ASCE*, Vol. 103, 489-499.

Ioannou, P. G. (1989) UM-CYCLONE discrete-event simulation system: user's guide, *Rep. UMCE-89-12,* Dept. of Civ. Engrg., Univ. of Michigan, U.S.A.

Martinez, J., and Ioannou, P. G. (1994) General purpose simulation with Stroboscope, *Proc. 1994 Winter Simulation Conf., IEEE*, Piscataway, N.J., U.S.A.

Portas, J., and Abourizk, S. (1997) Neural network model for estimating construction productivity, *J. Constr. Engrg. and Mgmt., ASCE*, 123(4), 399-410.

Smith, S. D. (1995) Productivity estimation of earthmoving operations using a discrete-event simulation model, *PhD Thesis*, University of Edinburgh, UK.

Smith, S. D., Osborne, J. R., and Forde, M.C. (1995) Analysis of earth-moving systems using discrete-event simulation, *J. Constr. Engrg. and Mgmt., ASCE*, 121(4), 388-396.

Sonmez, R., and Rowings, J. E. (1998) Construction labor productivity modeling with neural networks, *J. Constr. Engrg. and Mgmt., ASCE*, 124(6), 498-504.

CONCRETE PLACING PRODUCTIVITY USING A NOVEL NEURAL NETWORK DESIGN

Doug R. Forbes, Harry McGurnaghan, L. Darren Graham[1] and Simon D. Smith

School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN

For planning purposes an accurate estimate of the productivity for insitu concreting operations is desirable. Various methods have been investigated to model such productivity; this paper will consider a novel neural network approach in the search for increasingly accurate predictions. Neural networks are trained by providing them with a set of data from historic projects. Providing a larger training data set generally increases the accuracy of the output. In conventional neural network modelling, this training set is based upon the outcomes of a series of individual operations and thus is limited. The approach to be used here would expand this training set by considering the productivity of each individual concrete delivery used in each operation thus creating a much larger training set of data from which the neural network can learn. Further, by preparing the data in this way, an original architecture can be developed for the model. This architecture is a twin-loop nested recurrent network. The network is presented with each delivery for a given operation, and then with the productivity of the operation itself. Initial results are encouraging: the novel approach allows a much greater flexibility and a significant reduction in the sensitivity of changes to the network training functions.

Keywords: concreting, neural networks, planning, estimation, modelling

INTRODUCTION

To obtain an accurate estimate of concrete placing productivity, it is advantageous to create an analytical model that will allow such an estimate to be investigated to determine the response if variables are different. Neural networks (NNs) are a means of forming an approximation that have been used to model the concrete placing process by Sonmez and Rowings (1998) and Adeli and Wu (1998).

NNs learn from experience and consist of neurons, or processing elements, that are interconnected. The position and strength of the interconnection weights determine the predictive capabilities of the network. In training a NN it is these interconnections that are altered to best reproduce an estimate of the problem. Thus, training a NN is vital in efficient operation. In general, by providing a larger training set, more accurate solutions can be obtained.

This study examines the use of the standard NN architectures of feedforward (FF) and Elman networks. Following this, a novel network architecture, the *Twin Nested Recurrent Network* (TNRN) is developed which makes use of the concrete delivery process by providing a larger training data set for the network.

[1] Corresponding author: darren.graham@ed.ac.uk, 0131 650 7207

DATA ANALYSIS

Factors effecting the data

Many factors affect the concrete placing process. Crombie (1999) listed 44 separate factors that have an effect. Factors can be classified as quantitative or qualitative. Quantitative factors are numerical values. Qualitative factors are descriptions. Portas and AbouRizk (1997) applied a weighting to qualitative subjective factors depending on the influence of the factor. Rowings and Sonmez (1996) produced a NN model of the concreting operation using ten factors: quantities completed, job type, number of workers, temperature, humidity, precipitation, percentage labourer, percentage overtime, cumulative quantities and concrete pump.

The use of a NN to model this operation is due to its complex nature and the ability of the model to incorporate all the factors. However, the only variables that should be used in the model are those that can be established by the estimator, or the model user, before the project begins (Smith, 1999). Since the data for this model development was collected by others the data to be used was limited to that available.

Data Collection

It is important for a neural network to be presented with a set of data for training that goes right to the extremes of the problem domain (Flood and Kartam, 1994) to prevent the model extrapolating the results (Hecht-Nielsen, 1990).

While collecting the data, it is important to note the size of the sample. For this study, the sample size was 249. Rowings and Sonmez (1996) used a sample size of 112, whereas, AbouRizk et al. (2001) use a sample of 39. Arditi and Tokdemir (1999) used a sample size of only 12.

The data for the study was taken from observation. These observations were carried out by others. The contractors and the corresponding clients for the sources of original data are shown in Table 1.

Table 1: Sources of data

Contractor	Client
A	Scottish Water
B	Scottish Water
C	British Waterways
D	Highways Agency (HA)

The data for the Scottish Water projects relates to the construction of wastewater treatment works (WwTWs) and the associated tanks, chambers and bases. The data for British Waterways came from the construction of the Falkirk Wheel and for the Highways Agency the data was from the concrete pours involving bridges and underpasses during motorway construction.

Since the data was not recorded specifically for this study, there are certain factors that are not required, and certain factors that could be useful. The additional factors that could be useful will be discussed later. The main aspect of the data that was not required for this study was the identity of each individual truck mixer, and the corresponding cycle times. The remaining data allows a model to be created which can be used to experiment with neural networks in the concrete placing process.

From the data available, eight variables were used to model the productivity. This response was chosen as the output since cost and duration can be calculated from it. The eight factors were: month, location on site, average truck volume, total volume, average interarrival time, number of loads, number of accepted loads and number of rejected loads.

Delivery/Operation Data Analysis

In order to make full use of the data, a value of the productivity for each individual delivery was calculated. This was based upon the arrival time at the site of the delivery, the time at the end of the discharge and the volume of the delivery. Thus, the number of productivities equalled to the number of deliveries in an operation. The overall productivity of the operation was also calculated.

For the eight explanatory variables the data is clustered. The issue that arises is that the neural network will be trained with the same input sets for different values of productivity. The reason for this difference is that the factors that affect the productivity have not been measured, or are un-measurable. To resolve this problem, the standard deviations of the productivities for each operation were calculated. From this, a network could be trained to map to two outputs: the productivity and the expected standard deviation from this value. This data is then used to train and test the TNRN approach. By doing this a range of values can be given in a prediction, and is thus more trusted by users of the system than models that solely produce a point prediction value (Portas and AbouRizk, 1997; Lu et al., 2000; Shi, 1999)

Validation and Test Data

Test Data

The validation method was the train-and-test method (Twomey and Smith, 1997). From the 249 original operations, 210 were used in the development of the networks. This represents a 75:25 split of the original data. In order to ensure that a random sample was used for testing, a random number identifier was given to each of the data points. The data set was then sorted by the identifier and the first 158 points were used for training purposed. The remainder were used for testing.

Validation Set

The additional data from the HA project provided a validation set of 39 points. This data was not independent of the training data, having been taken from the same project, however it had not been used for training purposes (Shi, 1999).

THE NEURAL NETWORKS

Feedforward and Elman Networks

Feedforward and Elman networks were created to establish the ability of NNs to model the concrete placing process. The feedforward networks are a standard approach to NN modelling. The Elman network is a simple approach to a recurrent neural network (The MathWorks, 2003). Recurrent neural networks make use of a feedback loop. This loop allows the network to form temporal patterns, and hence provide better generalisations to the problem (Elman, 1990; Liu, 2001).

Network creation is a heuristic process for both these types of network (Hegazy et al. 1994; Adeli and Wu, 1998). It is necessary to alter the number of layers in the networks and the number of neurons within each layer, the training algorithms used,

and the activation functions at each layer (Hecht-Nielsen, 1989; The MathWorks, 2003).

From the work carried out using the train-and-test validation method, it was discovered that the networks were capable of predicting an average test MSE of 16 m^3/hr for the feedforward networks, and 22 m^3/hr for the Elman networks.

Twin Nested Recurrent Network Approach

The TNRN was specifically designed for this study and as far as could be determined similar architectures have not been previously developed. It combines feedforward networks with recurrent loops and makes use of the arrival of deliveries at a site to determine the productivity and the likely range the value lies within.

A diagram showing the network architecture is given in fig. 1. This approach uses two recurrent loops. For operation one, the first delivery is used to train the network and adjust the weights. This is then repeated for delivery 2, up to Q deliveries in operation 1. After each delivery the weights set by the previous delivery are used as the initial weight for the training. The network is then trained with the data for the first operation. The weights set by the delivery at this point are then used as the initial weights in operation 2, up to operation M. Once the data for operation M has trained the network, the weights set at this point are used for the final model.

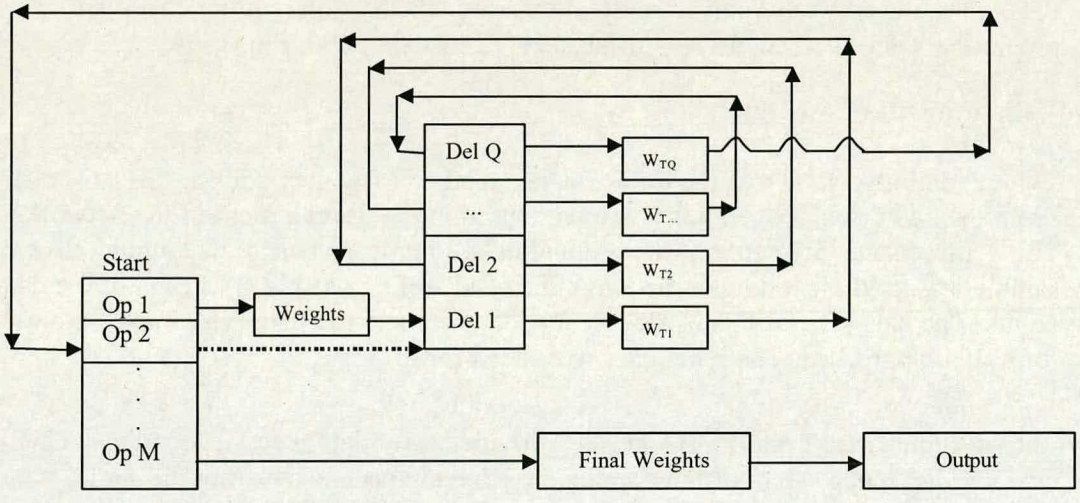


Figure 1: TNRN approach

This network was built with one hidden layer, and one output layer. Tests on the network showed that the output was independent of the number of neurons in each layer. It was also discovered that the network could not be trained without a linear activation function in the output layer.

From this, the network was created using 9 Matlab training algorithms and 3 combinations of activation function – Tan-sigmoid (T), log-sigmoid (L) and linear (P) in the first layer and linear in the second. The initial weights were set as the four combinations of 1 and 0. This created 108 networks. The networks used the same training set as the FF and Elman. The networks were trained with two outputs: productivity and standard deviation.

Results from TNRN

Training time for each network combination was approximately 100 minutes. This was not affected by the training algorithm or activation functions.

The results showed that for all four combinations of initial weights, for activation functions T-P and L-P, the productivity MSE was a constant value of $1060(m^3/hr)^2$ for all the training algorithms. However, for a P-P activation function, the MSE of the productivity varied from $611(m^3/hr)^2$ to $588,860,369(m^3/hr)^2$. The combination with the optimum value was:

- Initial weights: Input weights – 0; Hidden Layer Weights – 1
- Activation function combination: Linear-Linear
- Training algorithm: One Step Secant Algorithm

VALIDATION

Validation of TNRN Approach

The validation data results for the two outputs from this network are given in table 2 for the correlation co-efficient and the root mean squared proportional error, this is given for the productivity and standard deviation response. Figures 2 and 3 show the actual productivity and that modelled by the networks, for the productivity and standard deviation response respectively. As can be seen the relationship between the actual and the fitted data is very poor. All of the values in the table indicate that the two sets of data are not similar.

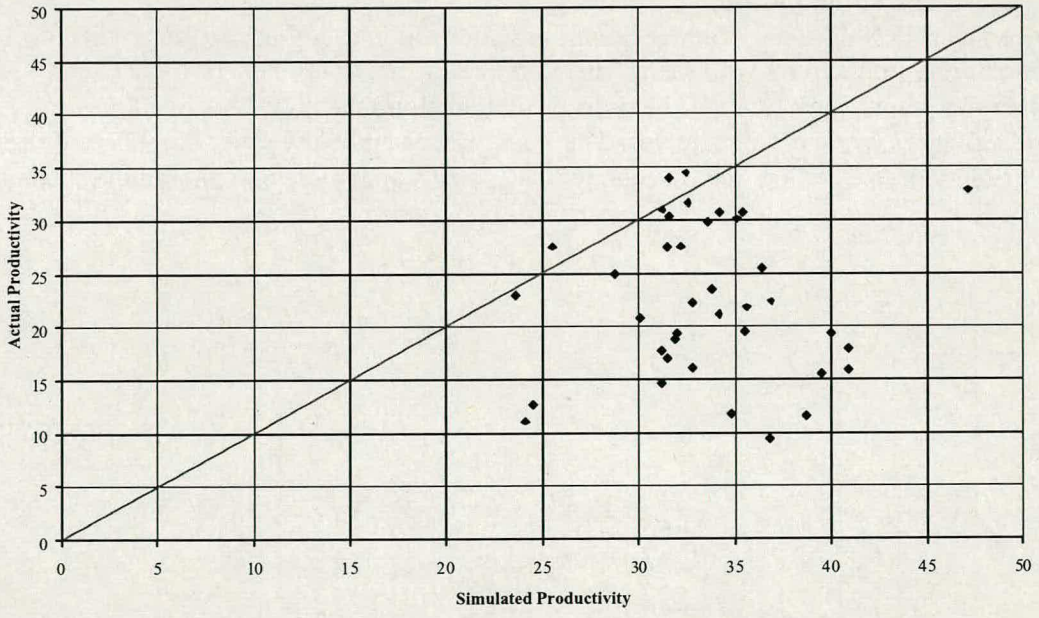


Figure 2: Simulated -v- Actual Productivity for Validation of TNRN

Table 2: Goodness-of-fit for validation data – TNRN

Response	Correlation Co-efficient, R^2	RMS Proportional
Productivity	0.016	0.967
Standard Deviation	0.042	1.043

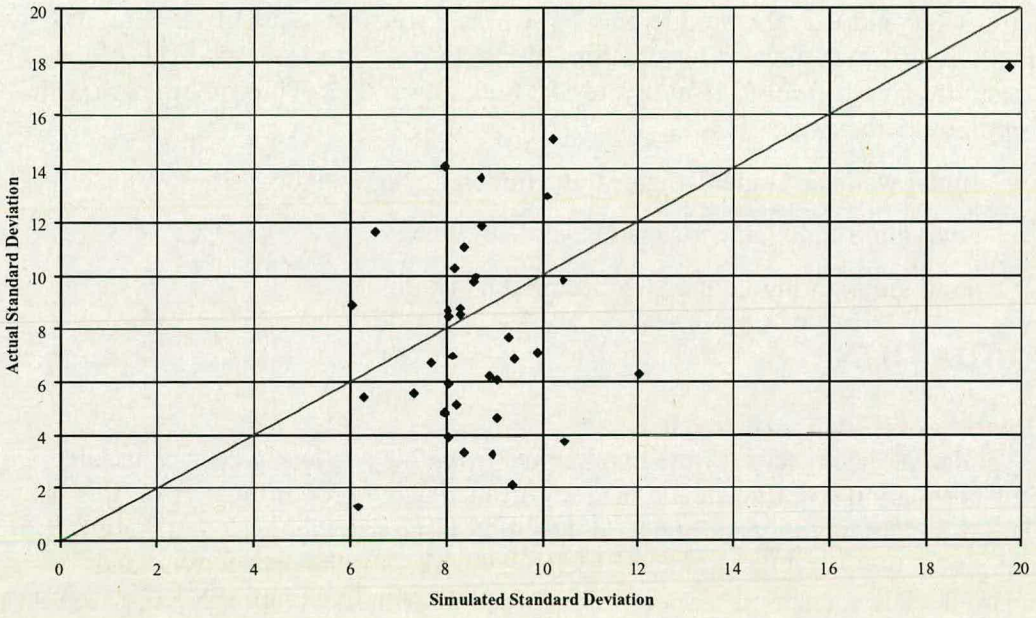


Figure 3: Simulated -v- Actual Standard Deviation for Validation of TNRN

Standard Deviation Output

Since the TNRN does not produce point predictions, it is useful to see the predicted value of the productivity with the predicted range. In fig. 4, for each of the 39 validation points, the actual productivity is plotted along the dots. The fitted productivity is marked with squares. The lines above and below show the range of the predicted value. The first tick on each indicates one standard deviation, and the second two.

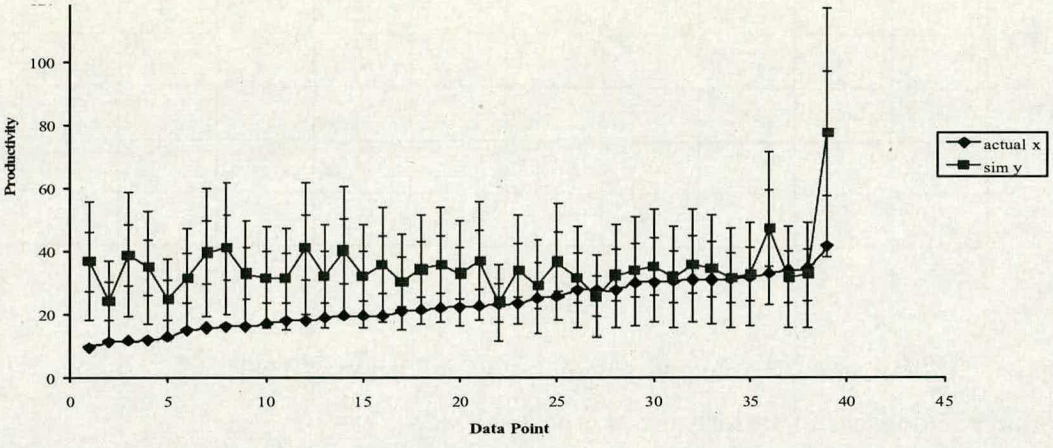


Figure 4: Actual and predicted outputs for TNRN, including standard deviation bars of predicted outputs

From fig. 4, it can be seen that for most of the predictions, the productivity is higher than required. However, all but four of the actual productivities lie within 2 standard deviations of the predicted value. It can also be seen in fig. 2 the use of the range of outputs from the TNRN. Better predictions are within the productivity range 22-38

m^3/hr. This is likely due to a larger amount of data in this range, and so the network can create a better generalisation pattern for these sets of variables.

Network Appraisal

On the face of it, the TNRN does not seem to provide good predictions to the problem. The generalisations that it provides tend to be larger than required, thus it would not be suitable for use as it does not give the same confidence as some of the FF and Elman networks. However the range of predictions does allow a value to be obtained. Two main advantages of this network are that it makes use of all the possible data within the concrete placing system, and provides a range of values as opposed to a point prediction.

DISCUSSION AND CONCLUSIONS

Data Discussion

The data that was used for this study was not ideal, since it had not been collected for the purposes of this study. Some of the factors were incomplete, and some were not measured at all. Eight explanatory variables were used in the neural network models. It is important to note that these explanatory variables are all measurable by an estimator before a project starts.

These eight are significantly shorter than the list produced by Crombie (1999), of 44 factors. It also did not include three of the factors listed by Rowings and Sonmez (1996) all of which related to the weather conditions. The weather cannot be predicted before the start of the project, but for the use of the models to predict productivity, had the factors been different, would be superior with weather included, and a comparison would be useful of a model with weather included and one without.

The list by Crombie includes many factors that are not measurable, thus the inclusion in an analytical model would not be appropriate. However, in this study, several improvements could have been made if a better knowledge of each individual operation had been available.

Type of Pour: The reference given to each type of pour (1-6) is all that could be ascertained based upon the data. However, this is not wholly accurate. For example, at the GPO Building in Edinburgh the pour was a simply supported floor slab. This did not reflect the pump configuration that involved four 90° bends, thus increasing the risk of blockages. The type of pour factor should be altered to reflect this.

Average Interarrival Time: This is an interesting factor to model, since it is dependent upon the delivery process. The interarrival time can be specified by the contractor before an operation, but it is unlikely that this will be followed exactly. For the data that was used in the model the specified interarrival times were not known. Thus, the actual interarrival times were use to construct the model. The arrival times of the deliveries are a factor in the wait times for the deliveries and the pump. This can have significant effects upon the production rate (Christian and Hachey, 1995). A more appropriate use of the data would have been to include a model with the number of late and early deliveries.

These two points show that the data is lacking. This is compounded by the analysis carried out to create the data set for the TNRN. For this, it was discovered that in each operation the levels of the eight factors were the same, but that each delivery had a different productivity based on the time it took to discharge. This meant that the eight

factors used as explanatory variables must not be exclusive and that there are other factors involved in the model, which are not included. These factors could have been measurable as discussed above, or un-measurable, and therefore could not have been successfully included in the model. However, based on the modelling results the factors used provided a good basis from which a model can be created.

The data is also limited by the ranges. For the month factor, the training and testing set only range from January to October. For the validation set, this is for all twelve months. The other factors have a similar range to each other. Thus, the networks can have problems in predicting data that is outwith the training.

Another limitation of the data that was used was that it only involved the concrete pouring system. This limits the data to one aspect of the system. It has been seen that the delivery process has an effect upon the wait times. Another important aspect of the system is the finishing of the concrete. This can vary in time, for example, a wall pour will have a much smaller area requiring finishing than a floor slab, and will thus take a shorter time. The data used in the model commenced with the time that the first delivery arrived, until the last delivery left. This made no account of any delay in the arrival of the first delivery, which could result in reduced productivity if the concreting operatives are waiting. However, this is a function of the delivery process, which can be modelled as a system in its own.

Feedforward and Elman Networks

The feedforward networks gave, on the whole better results than the Elman networks. This is possibly due to the Elman network being a more complex network architecture. This complex architecture means that the network is more susceptible to fluctuations in the data, and missing variables. It has already been noted that there are missing parts in the data that could make the Elman network operate more effectively. One of the major drawbacks of these networks is the point prediction that is made, that provides no confidence intervals to guide a user.

TNRN Recurrent Network

The network was based upon the research that has been carried out using a hybrid collection of standard networks, connected in novel ways to form different architectures (Liu, 2001; de Jesus et al., 2001; Ninomiya and Sasaki, 2002).

In the training and testing of the networks, it was discovered that the TNRN was not affected by the number of hidden neurons. This is likely to be due to the small amount of data that was used in each presentation. This is a major advantage, since standard Elman and FF networks require heuristic methods to define the number of neurons, which is time-consuming and cumbersome.

The transfer functions and training algorithms (TAs) used in the development of this model showed results that were not expected. Firstly, the transfer function required a linear function in the output layer, otherwise the network would not train. Thus, with a linear transfer in the output layer, when the hidden layer function was tansig or logsig, the values of MSE for both outputs were constant for all TAs. With a linear transfer function in the hidden layer, however, the network produced a different result for each TA. The range of these was much larger than those of the standard networks.

The validation results from this network show poorer values than that of the standard approaches. The correlation co-efficient of the data is very poor for both the productivity and the standard deviation responses. However, the aim of the TNRN is

not to produce a network that creates only a point prediction, but a network that will give a range of values for the given conditions. Most of the actual values lie within the two standard deviation range. However, from this plot, there is a trend for the network to be unable to make accurate predictions at the lower end of the productivity range. This is likely because training data does not extend fully to this area.

Overall, the TNRN network is likely to produce less accurate results, for similar reasons to the Elman NN. That is, that as the network architecture is more complex a better quality of data is required for training, as the network is more sensitive to fluctuations or errors. The TNRN provides promising results, and further development would be advantageous, perhaps with data from simulation to provide a larger data set.

CONCLUSIONS

From the study carried out, the following conclusions can be reached:

FF networks were shown to have the most accurate solutions. These NNs, however, had the disadvantage of taking a long time to develop, and only providing a point prediction.

The more complex the model created, the more important it is to use a set of training data that is of a high quality. Thus, for the Elman and TNRN approach, the errors in the predictions can be attributed to the limitations of the data. If more appropriate data is available for training, it may be possible to obtain better predictions with these methods.

The TNRN produces promising results that could be applied to other forms of cyclic operations with repeated events, such as earthmoving, steel erection or another delivery process.

Neural network modelling is a heuristic tool that can be time-consuming to define the most appropriate solutions. The TNRN, if developed to obtain more accurate solutions, would be advantageous since it cuts down on the number of component variables. Thus, the network development at this stage is not solely concerned with the end-user, but also with the developers. More work requires to be done in this area, to examine and define the pitfalls for modellers using NNs in modelling construction processes.

REFERENCES

AbouRizk, S, Knowles, P & Hermann, U R (2001) Estimating Labor Production Rates for Industrial Construction Activities. *Journal of Construction Engineering and Management*, **127**(6), 502-511

Adeli, Hojjat & Wu, Mingyang (1998) Regularization Neural Network for Construction Cost Estimation. *Journal of Construction Engineering and Management*, **124**(1), 18-24

Arditi, David & Tokdemir, Onur Behzat (1999) Comparison for Case-Based Reasoning and Artificial Neural Networks. *Journal of Computing in Civil Engineering*, **13**(3), 162-169

Christian, John & Hachey, Daniel (1995) Effects of Delay Times on Production Rates in Construction. *Journal of Construction Engineering and Management*, **121**(1), 20-26

Crombie, R (1999) *Efficiency, Quality and Waste in Concrete Operations.* BEng Thesis, University of Edinburgh

De Jesús, Orlando, Horn, Jason M & Hagan, Martin T (2001) Analysis of Recurrent Network Training and Suggestions for Improvements, *In: Proceedings of International Joint Conference on Neural Networks, 15-19 July 2001,* **4**, 2632-2637

Elman, Jeffrey L. (1990) Finding Structure in Time. *Cognitive Science*, **14**, 179-211

Flood, Ian & Kartam, Nabil (1994) Neural Networks in Civil Engineering I: Principles and Understanding. *Journal of Computing in Civil Engineering*, **8**(2), 131-148

Hecht-Nielsen, Robert (1990) *Neurocomputing*, Wokingham: Addison-Wesley

Hegazy T, Moselhi, O & Fazio, P (1994) Developing Practical Neural Network Application Using Backpropagation. *Journal of Microcomputers in Civil Engineering,* **9**(2), 145-159

Liu, Derong (2001) Open-Loop Training of Recurrent Neural Networks for Nonlinear Dynamical System Identification. *In: Proceedings of International Joint Conference on Neural Networks, 15-19 July 2001*, **2**, 1215-1220

Lu, Ming, AbouRizk, S M & Hermann, Ulrich H (2000) Estimating Labor Productivity Using Probability Inference Neural Network. *Journal of Computing in Civil Engineering*, **14**(4), 241-248

Ninomiya, Hiroshi & Sasaki, Ayako (2002) A Study on Generalization ability of 3-Layer Recurrent Neural Networks. *In: Proceedings of the 2002 International Joint Conference on Neural Networks, 12-17 May 2002*, **2,** 1063-1068

Portas, Jason & AbouRizk, Simaan (1997) Neural Network Model for Estimating Construction Productivity. *Journal of Construction Engineering and Management,* **123**(4), 399-410

Rowings, James E & Sonmez, Rifat (1996) Labor Productivity Modeling With Neural Networks, *In: AACE International Transactions, PROD.01*, Morgantown, WV, USA, 1-3

Shi, Jonathan Jingsheng (1999) A neural network based system for predicting earthmoving production. *Construction management and Economics*, **17**(4), 463-471

Smith, Simon D (1999), Earthmoving Productivity Estimation Using Linear Regression Techniques. *Journal of Construction Engineering and Management,* **125**(3), 133-141

Sonmez, Rifat & Rowings, James E (1998) Construction Labour Productivity Modeling with Neural Networks. *Journal of Construction Engineering and Management,* **124**(6), 498-504

The MathWorks (2003) *Neural Network Toolbox for use with MATLAB®: User Guide* [online]. Massachusetts: The MathWorks. Available from : http://www.mathworks.com/access/helpdesk/help/pdf_doc/nnet/nnet.pdf [Accessed 28 October 2003], pp. xviii, 5-1 – 5-50, 9-1 – 9-7

Twomey, Janet M, & Smith, Alice E (1997) Validation and Verification. *In:* Kartam, Nabil, Flood, Ian & Garrett, James H Jr, (eds.) *Artificial Neural Networks for Civil Engineers: Fundamentals and Applications.* New York: ASCE, 44-64

A Model to Facilitate the Optimization of the Concrete Supply Process

L. Darren Graham[1] and Simon D. Smith[2]

[1]Institute for Infrastructure and Environment, The University of Edinburgh, AGB Building, EH9 3JN, UK; PH (0044) 131-650-7207; FAX (0044) 131-650-6781; email: darren.graham@ed.ac.uk (corresponding author)
[2]Institute for Infrastructure and Environment, The University of Edinburgh, AGB Building, EH9 3JN, UK; PH (0044) 131-650-7159; FAX (0044) 131-650-6781; email: simon.smith@ed.ac.uk

Abstract
The current method of planning a concreting operation is for a construction practitioner to make an intuitive estimate about the quantity of resources (people and plant) required in a particular situation. This has been found to produce inefficient operational performance, which is unacceptable in the contemporary climate where construction contractors receive in the region of 3% of the total project budget as profit. An alternative method of providing such estimates is presented here, which consists of using a computer model, *CBRSim*, powered by a combination of artificial intelligence and statistical methods, to determine the likely achievable performance in the ready mixed concrete (RMC) supply process. It was found that *CBRSim*, can recreate past concreting operations with a high degree of accuracy, and thus the output from the model carbe trusted to a certain degree.

In this paper, a previously unpublished method to attain an optimum performance in the RMC supply process using *CBRSim* is presented. This entails: undertaking experiments (using *CBRSim*) to examine the effect of varying the number of truck mixers on performance; from these experimental results an identification of the optimum performance can be made, and the corresponding details regarding the quantities of resources required for an operation can be relayed to a supplier.

Introduction

Ready mixed concrete (RMC) is an important material in contemporary construction. As such, the management of the supply process of RMC should be effective enough to ensure efficiency. Unfortunately this is not the case (Smith, 1998a): the current method of choosing the quantities of resources (e.g. plant and people) required for a particular operation is an intuitive estimate. This technique is unable to take into account the stochastic nature of the process and this has been found to result in site-

based resources (concrete pump, people, etc.) remaining idle for 12% of the overall duration at a cost of 14% of the total (Graham et al., 2005). These levels of waste in the RMC supply process are comparatively small to the average waste witnessed in construction projects of roughly 50%. However, the waste incurred in using RMC is undoubtedly a contributing factor to the high waste levels witnessed in construction projects. Thus, any reduction in the amount of waste observed in the RMC supply process is a step in the right direction to reducing the waste observed in whole construction projects.

There is a need for better selection of the resources required in a particular concreting operation. This selection process can be aided significantly by a computer model, which graphically displays the likely performance (productivity, utilization of on-site resources, etc.) achievable in an RMC supply process, given a set of conditions. A computer model, *CBRSim*, has been developed using VBA in Microsoft Excel (Graham et al. (2004)), to perform this very task. *CBRSim* consists of two modules: a case-based reasoning (CBR) model that determines the statistical characteristics of a particular operation; and a simulation model which uses these statistical characteristics to create an accurate graphical display of the operational performance given differing quantities of resources. A typical graphical display produced by *CBRSim* is shown in figure 1. From such a display, the selection of the resources that produce the optimum operational performance is possible.

Here, the *CBRSim* model, which is based on real construction data, is presented and validated. Subsequently, a case study is shown to demonstrate the use of *CBRSim* in the optimization of the RMC supply process. Prior to outlining the model, this paper will look briefly at the process of RMC supply.

The RMC Supply Process

In previous research, concreting operations have been separated into two processes: batching and delivery (for example, Sawhney et al., 1999); and pumping and placing (Dunlop and Smith, 2000). In this study, the RMC supply process being modeled incorporates all of the above elements and in an ideal situation: a truck mixer is filled with RMC at a batching plant; then travels to a construction site; upon arrival, the truck mixer moves into position at a concrete pump and discharges its load; any residue is cleaned from the truck mixer in a washout area; finally, the truck mixer returns to the batching plant, completing the cycle. In reality, the events that occur within the system (e.g. time that truck mixers arrive on site) take place at irregular intervals (Smith, 1998b). Queuing of truck mixers could occur at the batching plant, pump and washout area, as it is possible that a truck mixer will not arrive at these places before the departure of the previous truck mixer – this is known here as over utilization of the on-site resources (see figure 1); where the on-site resources are the plant, pump and washout area. This situation can result in a steady supply of RMC to the concrete pump often resulting in a high rate of process productivity. However, an RMC supplier adds a charge for every minute that a truck mixer spends on site, over an agreed rate. Thus, with truck mixers sitting idle on site, this situation can prove to

be expensive. Conversely, it is also possible for a gap in the supply of truck mixers to occur, resulting in a concrete pump, or washout area that is idle – this is known as under utilization of the on-site resources (see figure 1). This situation often results in a lengthening of the process and a lower rate of process productivity than that witnessed in the case of resource over utilization. As the plant and people involved in an operation are priced over time, the longer an operation lasts the more it costs. Thus, when the on-site resources are being under utilized the cost of an operation can rise substantially.

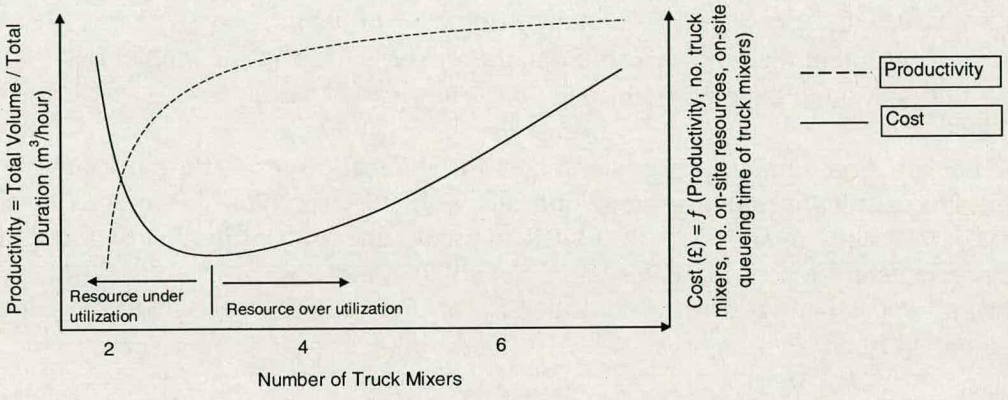


Figure 1.Typical performance of the RMC supply process (obtained from *CBRSim*)

The stochastic nature of RMC supply can be recreated using probability distributions to represent the following times that exist within the process: load truck mixer (batch) time; travel time from batching plant to construction site (travel to site) time; time to position truck mixer next to concrete pump (position); discharge (pump) time; time to clean the truck mixer (washout); travel time from site to batching plant (travel from site). Samples of each of these times are provided by the probability distributions, which are used to recreate a typical cycle of the process. Through the simulation of multiple cycles, the performance of a particular operation can be estimated.

The CBRSim Model

To ensure that the model is realistic and accurate it must be based upon data collected from real construction projects. In this study, a total of 225 concreting operations were observed on a wastewater treatment works project. Additionally, 20 operations observed on a highway viaduct project were not used in model development, but were retained for use in model validation.

Simulation has been used extensively in the past to model construction operations and project performance (e.g. Tommelein, 1998; Tommelein et al., 1999; Horman, 2001). One major drawback to simulation is that the user is required to define the probability distributions and parameters that represent the times that exist within a

system. Here this difficulty is removed by using a CBR model to provide an input (the distributions, etc.) to a simulation model.

CBR Module

CBR is the process reasoning and learning by storing cases – records of specific prior reasoning episodes, which are construction operations in this model – and retrieving and adapting them to aid new problem solving or interpretation in similar situations (Kolodner, 1993). Aamodt and Plaza (1994) described the CBR process as cyclical, comprising of the following:

- Retrieval of the most similar case(s) from the case base
- Reuse of these case(s) to attempt to solve the problem
- Revision of the proposed solution, if necessary, to solve the problem
- Retention of the modified solution as a new case
-

The construction domain is a growing area for applications of CBR and it has been used successfully in solving prediction and estimation problems (Morcous et al., 2002). This study intends to utilize CBR to predict the probability distributions and their parameters that best represent a particular concreting operation. *CBRSim* is based upon the cyclical process described by Aamodt and Plaza and its methodology is shown in figure 2.

The *CBRSim* model was developed in two stages: case base design and a case retrieval mechanism.

(a) Case base design

The case base is the storage area for prior cases (examples of construction operations in this model). Cases must be parameterized, typically by being divided into a series of distinct features by means of which their similarity to other cases can be judged. In *CBRSim*, the case base initially stored 225 cases, with all the cases sharing the same record structure of 4 input and 6 output features. The input features were chosen because they can easily be determined in practice and consist of: the number of trucks, pour volume, month of pour and the type of pour. Each input feature is assigned an index to aid the retrieval process (see figure 3). The output features are the probability distributions and parameters that best represent the six times that exist within the process (mentioned above).

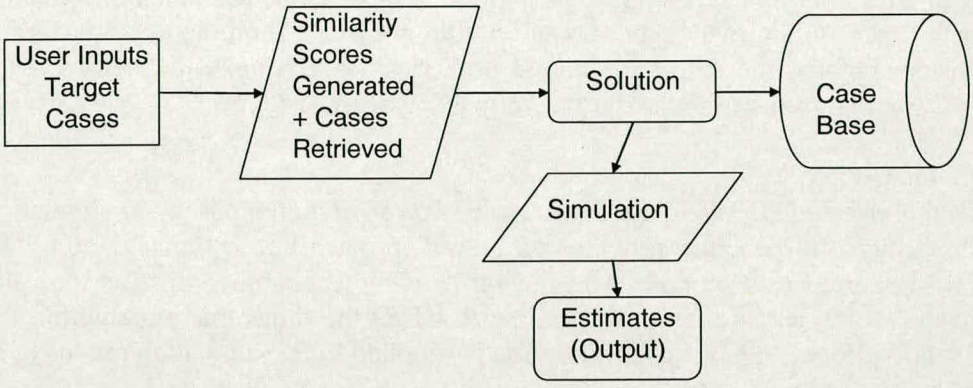


Figure 2. *CBRSim* model methodology

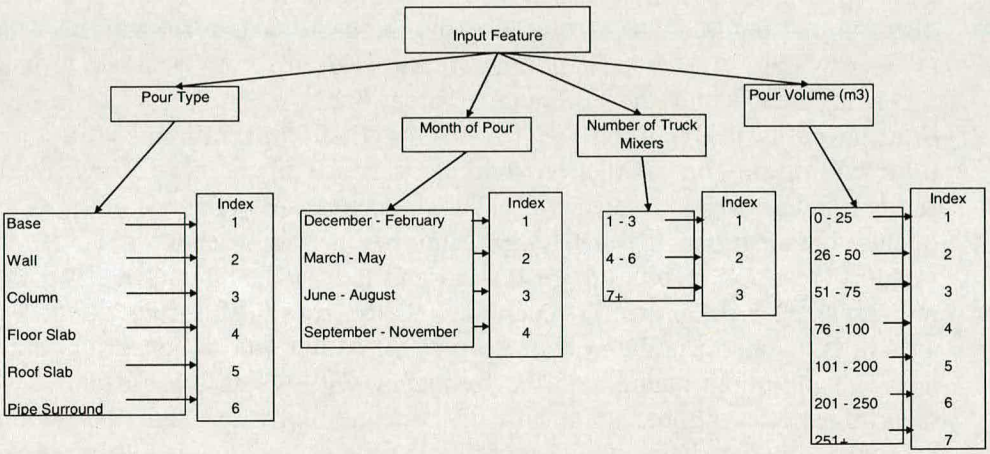


Figure 3. Input feature indexing system used in *CBRSim*

(b) Case retrieval mechanism

A CBR model derives its power from its ability to retrieve relevant cases quickly and accurately from its case base (Arditi and Tokdemir, 1999). To perform the retrieval process, *CBRSim* measures the similarity between a user specified target case and the input feature indices of each stored case. This similarity measurement is in percentage form and is based upon Euclidean distance measurements, as detailed in Hand et al. (2001). If the similarity between the target case and a stored case is greater than 75%, the stored case is retrieved from the case base and is placed in a temporary storage area. Once all the relevant cases (> 75% similarity) have been place in this temporary space, the user is permitted to view them. The user is then required to select a case, and hence a set of probability distributions and parameters, based only on a percentage measurement – removing the need for a user to have a knowledge of statistics. If the input features of the selected and target cases are non-

identical, *CBRSim* retains the target case input features and the statistical information as a new case – this provides the model with an ability to learn, and completes the CBR cycle. Finally, the case base should be updated over time with the details of new construction projects, to bolster the knowledge of *CBRSim*.

Simulation Module

As stated above in *CBRSim*, the CBR module is providing an input to the simulation module, which subsequently provides an output (productivity estimates, etc.). The input to the simulation module is in the form of the probability distributions and parameters which best represent (according to CBR) the times that exist within the RMC supply process. Once this information is supplied to the simulation module two actions are performed:

1. **Random samples (variates) are generated from the probability distributions by a random variate generator**. The method used is to transform a uniformly distributed random number (0 to 1) through an algorithm, into a series of numbers that correspond to the particular distribution required. 1000 of these numbers are averaged and it is this average which forms a variate. One variate is used to represent each of the times that exist within the system (e.g. batch time).
2. **Simulation is undertaken for the user-defined operational set-up (e.g. pour volume).** The simulation module is based upon three FIFO single-server queues, working in series. These queues are representative of real queues occurring at the batching plant, pump and washout area. In this system, the truck mixers are customers and the batcher, pump and washout are servers. When an event occurs, such as the arrival of a truck on site, the system is updated. As the system evolves over time, trucks depart and others arrive at the pump queue, and the frequency with which these events occur determines the number of loads of concrete delivered. As the module proceeds, other statistics such as pump idle time or the truck wait time in each queue may be recorded. From these times, values of interest such as pump utilization and truck utilization can be calculated. Simulation is run from an event list, which contains the next time when each type of event will occur. This time is a sample from the chosen distribution, produced by the random variate generator. For instance, the arrival time of a truck on site is generated by addition of the time of departure from the batching plant plus a sample of the travel time distribution. Finally, simulation will be completed when the user-specified concrete pour volume has been placed. The output of the simulation module is presented in two graphs: how the utilization of truck mixers and utilization of on-site resources vary with the number of truck mixers (e.g. figure 5); how productivity and cost vary with the number of truck mixers (e.g. figure 6).

Model Validation

Validation is the process of determining whether a model is an accurate representation of a system (or process). *CBRSim* is intended to represent the RMC

supply process, not replace it, in order to allow predictions of the process performance to be made. Thus, the model's predictions of past operational performance and the observed performance may be compared to ascertain if the model is valid. 20 concreting operations (validation observations) were used in this study, and only the predicted and observed process productivities were compared. The results of the model validation are shown in figure 4 and it is clear that *CBRSim* is capable of reproducing the performance of the RMC supply process to a high degree of accuracy, as the fit line rarely strays from the line indicating 100% accuracy (1:1).

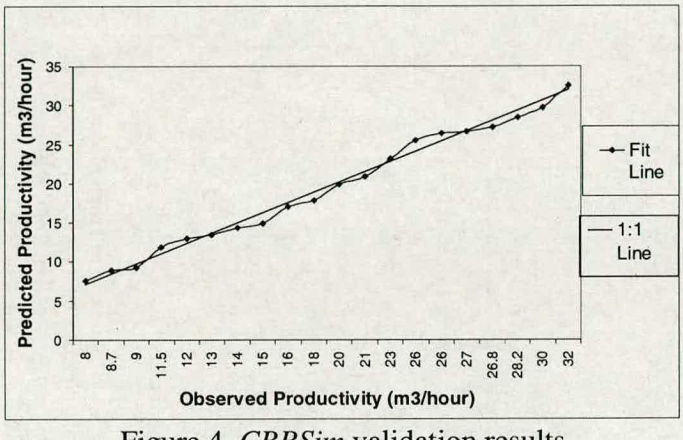


Figure 4. *CBRSim* validation results

Optimization of the RMC supply process

CBRSim is capable of experimenting with the effects of varying the number of truck mixers used in a particular operation. The results of these experiments are provided in two graphs (see figures 5 and 6 for the case study results). It is these outputs which are used to optimize the performance of the RMC supply process. The steps are simple:

1. Determine upon the number of truck mixers which result in maximum utilization of both the truck mixers and the on-site resources (otherwise known as the point of balance).
2. Confirm that the number of truck mixers chosen from point 1 provide the optimum value (maximum productivity, minimum cost). It is usually the case that this is so, as the cost increases rapidly with either an over or under utilization of on-site resources. However, there may be exceptions to this rule and it is prudent to perform this check.
3. Relate the chosen number of truck mixers to the concrete supplier.

The following is a case-study of the optimization of the RMC supply process:
Case details. The case in point is a 210 m^3 base pour, carried out in May. The number of truck mixers to be modeled is 3 to 10. *CBRSim* recreated the process performance for each number of truck mixers, producing 8 results to compare in this instance.

The three steps details above are then carried out:

1. One of the plots from *CBRSim* shows the utilization of on-site resources against the utilization of the truck mixers as the number of truck mixers is varied (shown in figure 5). This allows the point of balance between these two measurements to be identified – this is the point on the plot where the two lines intersect – and corresponds to using 6 truck mixers in this particular operation.

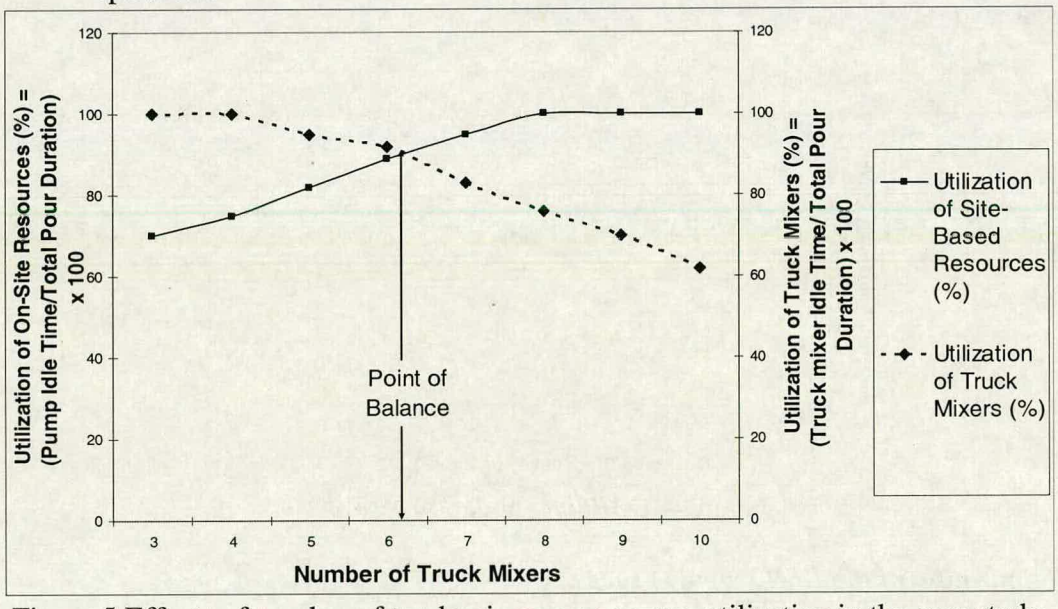


Figure 5.Effects of number of truck mixers on resource utilization in the case study.

2. After determining the point of balance between the utilization of on-site resources and truck mixers, it is necessary to confirm that the number of truck mixers which produced this balance also produces the optimum process performance. It is usually the case that this is so - cost increases rapidly with either an over or under utilization of on-site resources. The other output plot produced by *CBRSim* (shown in figure 6) must now be used. This plot shows that in this case 6 truck mixers (the point of balance) produces value (maximizing productivity while minimizing cost), and consequently has optimized the RMC supply process.
3. When ordering this particular batch of RMC the construction practitioner should request that 6 truck mixers should be utilized in the particular operation. If this request is made with a large enough lead-time, the RMC supplier should be able to arrange that the required number of truck mixers is made available for a particular operation.

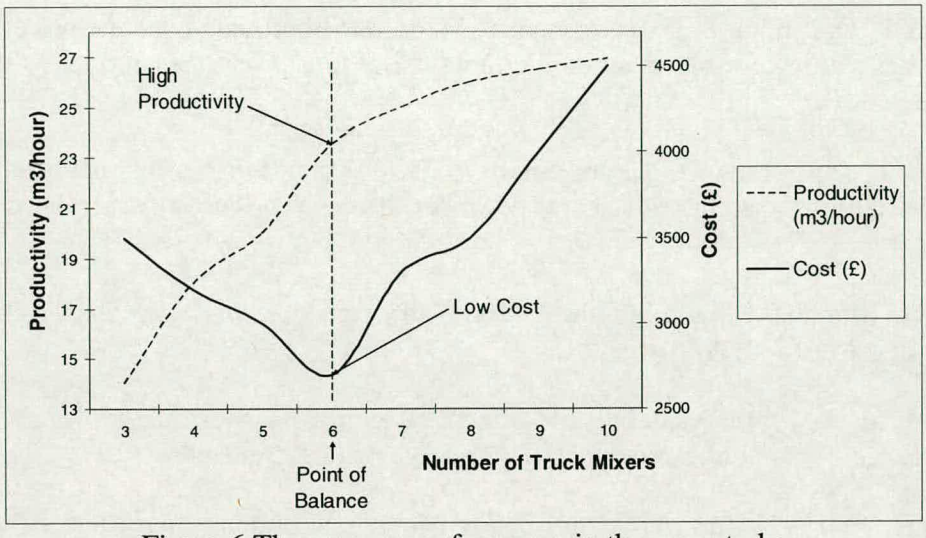


Figure 6.The process performance in the case study.

Conclusions

The RMC supply process was accurately modeled using a combination of CBR and discrete-event simulation. This model, *CBRSim* possessed the capability to allow experiments to be performed to examine the effect that varying the number of truck mixers used in an operation, had on process performance. The model's outputs from these experiments are in graphical form, permitting a simple identification of the quantity of resources that produces an optimum process performance. It was found that the optimum process performance provides 'value' to the contractor – it is a compromise between operational cost and duration. It was also found that this optimum often coincides with the number of truck mixers that result in a balance between the rates of utilization of on-site resources (pump and washout) and the truck mixers themselves. This balance ensures that the pump, washout and truck mixers collectively remain idle for the minimum possible amount of time – which has a further benefit of reducing the space required to perform a concreting operation, due to the minor amount of queuing of truck mixers.

References

Aamodt, A., and Plaza, E. (1994) Case-based reasoning: foundational issues, methodological variations, and systems approaches, *AI Communications*, 7(i), 39-59.

Arditi, D., and Tokdemir, O.B. (1999) Comparison of case-based reasoning and artificial neural networks, *J. Comp. In Civil. Engrg., ASCE*, 13(3), 162-169.

Dunlop, P., and Smith, S. D. (2000) Stochastic modelling of concrete operations, *In:* Akintoye, A. (Ed.), *16th Annual ARCOM Conference*, 6-8 September 2000, Glasgow Caledonian University, Vol. 2, 483-492.

Graham, L. D., Smith, S. D., and Crapper, M. (2004) Improving concrete placement simulation with a case based reasoning input, *Civil Engrg. and Environ. Syst.*, Vol. 21(2), 137-150.

Graham, L. D., Forbes, D. R., and Smith, S. D. (200X) Modeling the ready mixed concrete delivery system with neural networks, *Automation in Construction,* under review.

Hand, D., Mannila, H., and Smyth, P. (2001) *Principles of data mining,* MIT Press, Cambridge, Mass., USA.

Horman, M. J. (2001) Modeling the effects of lean capacity strategies on project performance, *9th Annual Conference, IGLC*, August 2001, Singapore.

Kolodner, J. (1993) *Case-based reasoning,* Morgan Kaufmann, San Mateo, Calif., USA.

Morcous, H., Rivard, A. M., and Hanna, A. M. (2002) Case-based reasoning system for modeling infrastructure deterioration, *J. Comp. in Civil. Engrg., ASCE*, 16(2), 104-114.

Sawhney, A., Abudayyeh, O., and Chaitavatputtiporn, T. (1999) Modeling and analysis of concrete production plant using petri nets, *J. Comp. in Civil. Engrg., ASCE*, 13(3), 178-186.

Smith, S. D. (1998a) Concrete placing analysis using discrete-event simulation, *Proc. Institution of Civil Engineers, Structures and Buildings*, 128 (4), 351-358.

Smith, S. D. (1998b) Stochastic analysis of cyclic construction processes, *in:* Hughes, W. (Ed.), *14th Annual ARCOM Conference*, 9-11 September 1998, University of Reading, Vol. 1, 127-136.

Tommelein, I. D. (1998) Pull-driven scheduling for pipe-spool installation: simulation of lean construction technique, *J. Constr. Enrgrg. and Mgmt., ASCE,* 124(4), 279-288.

Tommelein, I. D., Riley, D. R., and Howell, G. A. (1999) Parade game: impact of work flow variability on trade performance, *J. Constr. Enrgrg. and Mgmt., ASCE,* 125(5), 304-310.

COST OF CONCRETE PLACEMENT FOR CONTRACTORS

Darren Graham[1], Simon D. Smith[2] and Iris D. Tommelein[3]

[1, 2] *School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, UK*
[3] *Department of Civil and Environmental Engineering, The University of California at Berkeley, Berkeley, 94720, USA*

RMC is an important construction material and the cost of placing this material is significant for a contractor. Previous research studies to minimise this cost have centred upon the cost of plant and labour, without consideration of the other costs that comprise the RMC system. The aim of this research is to define the costs that should be reduced to minimise the contractor cost in the RMC system and identify a method(s) of achieving this minimisation.

The costs and the contributor(s) to these costs were identified through consultation with industrial practitioners. A quantitative analysis was performed on a data set collected from UK construction projects to determine the significance of each cost and the contributor(s) to each cost. The significant costs were the RMC material, plant hire, idleness of plant and on-site labour (placing team) and truck mixer surcharges. These costs should be reduced.

A systemic cost reduction would include all of the significant costs identified and produce a larger reduction in the cost of RMC placement for contractors than focusing on plant and labour costs alone. Such a systemic would require a fundamental re-management of the RMC system such as a reconfiguration of the resource supply chain coupled with production control methods aimed at reducing plant and labour costs.

Keywords: cost, materials management, production management, ready-mixed concrete.

INTRODUCTION

Ready mixed concrete (RMC) is an important material for contemporary construction, as it is used in a variety of structures, for example in motorway viaducts, building foundations and wastewater treatment plants. The global construction industry is a major consumer of RMC, and the trend that production rates of the material are on the increase (Eurostat, 2005) signifies a growing reliance by the construction industry on the material. It is thus of concern that RMC is not being manufactured, transported and placed with a high level of efficiency and at a minimum cost to the supplier, contractor and the project client (Halpin and Woodhead, 1976; Camillo, 1996; Smith, 1998; Zayed and Halpin, 2001).

It is clear from the dates of the above publications that this problem has not only been realised in the recent past; rather it has been a continuing problem for decades. However, as stated above there is an increasing reliance on RMC as a construction material and this is the catalyst for this research, which aims to define the sources of

[1] Tel: 0131 650 7159, e-mail: simon.smith@ed.ac.uk

contractor cost in RMC placement and outline a method(s) of reducing these costs. The study is presented in two parts, with this paper defining the sources of contractor cost in RMC placement and another paper presented by Graham et al. (2005) at ARCOM 2005 entitled 'Cost reduction through vertical integration of the in-situ concrete supply chain' outlining a cost reduction method(s).

This study concerns the RMC system when an external batch plant is acting as RMC supplier and a contractor is placing the RMC using a concrete pump – a common situation. Additionally, this study is ongoing and here it is limited to considering the costs incurred by contractors in placing concrete – a future study shall address the issue of the costs for the RMC supplier.

Previous research has suggested that a minimisation of contractor cost for placing RMC can be achieved through optimising the allocation of the required resources (truck mixers, plant, people, etc.) and thereby reducing the cost of labour and plant (for example Zayed and Halpin, 2001). This area has been focused upon because the allocation of resources is to a large degree within the control of the contractor and thus these cost reduction methods could be implemented readily. While optimising the allocation of resources would result in a decrease in the cost of RMC placing, it is felt that this reductionist view is not identifying all of the costs involved in the placement of RMC. Moreover, the cost of plant and labour may not be the most significant cost that a contractor must bear in the RMC system and thus there could be other costs that should be reduced in addition to those of plant and labour. To this end, an analysis of the data collected from real RMC pours has been made to:

- Identify and quantify the costs that exist in the RMC system for a contractor;
- Expose the contributor(s) to these costs and assess the significance of each.
- Define areas of focus for contractor cost reduction strategies.

In this article, firstly the details of the sources of the project data are presented. Next, the contractor costs within the RMC system are identified and subsequently quantified through an analysis of the project data. A comparison of this cost analysis is made with existing cost information provided by RS Means. The contributor(s) to the quantified costs are then examined. Finally, areas of focus for contractor cost reduction strategies are defined based upon the analysis performed in this study.

PROJECT DATA

The project data was collected during a time and motion study of the construction of six wastewater treatment plants in two different projects, performed by two different contractors in Scotland, UK. Project one (4 plants) had two suppliers of RMC – located at 5 and 18 miles from the project site. A total of 184 RMC pours and 14,900m^3 of placed RMC were observed in this project. Project two (2 plants) had two suppliers – located at 8 and 16.5 miles from the project site. 48 RMC pours and a total volume of 6, 200m^3 of placed RMC were observed in this project.

The information collected included: date, location (on the structure) and type of pour; weather (qualitative); truck mixer number; time batching started and finished; travel time for truck mixer; time of arrival of truck mixer on project; time that discharge began; finish time of discharge; time washout begins; time washout finishes; leave site time; travel time; time of arrival at batch plant; rejected loads; time lost due to rejected loads; time lost due to pump breakdowns; wait period of truck mixers (if any); time to position truck mixers; idle time of plant and labour due to a gap in the

supply of RMC; time truck mixers spend on project site; pour duration; no. labour force in each role (e.g. plant operator); and pour volume.

For generic conclusions to be drawn from this research, it is important to ensure that the project data is evenly and well-distributed. A measure of this can be made by presenting the collected data in a histogram. In this case, the data should be well distributed in pour volume (Anson et al., 2002) and productivity (Dunlop and Smith, 2003) – two key variables of the system. A histogram of the pour volumes, with the productivities of the project data is shown in figure 1.

The histogram in figure 1 is skewed to the right – there are more pours with a volume below 150m^3 than above. When the average productivity of each of the volume categories is plotted (as in figure 1) a significant difference is shown between those pours below 80m^3 and those above; pours with more than 80m^3 volume have a higher rate of productivity. To promote generic findings from this study the project data was separated into two distinct groups, those 80m^3 or below (small pours) and those above this volume (large pours). This is because each of these groups is more evenly distributed than the data set as a whole. The small pours observed in this study totalled 6, 000m^3 performed over a four-month period.

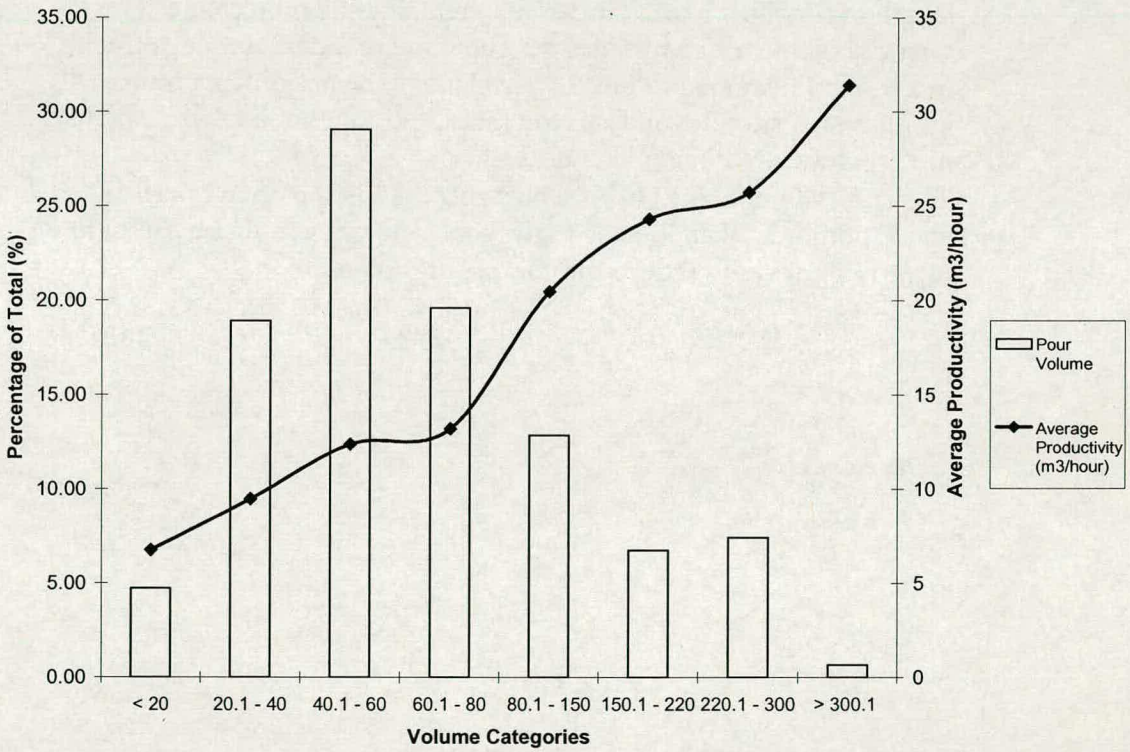


Figure 1: Project data pour volume histogram

IDENTIFICATION OF CONTRACTOR COSTS

The authors collaborated with several managers of both RMC batch plants (in USA and UK) and contractors (in UK) to identify some of the costs in the RMC system. These costs are presented in an Ishikawa or fishbone diagram (see figure 2). Note that this list is not thought to be complete. The costs for a contractor are as follows:

- RMC material costs. The price quoted by a RMC supplier absorbs a number of the supplier's costs, including raw materials prices, overheads, transportation costs and suppliers profits. These prices can vary significantly according to the design mix, but is typically in the range £30/m^3 to £50/m^3.
- Plant hire costs. The on-site plant required in the RMC system differs from pour to pour. However, in the studied projects there was some standard plant used in each pour, such as a concrete pump, vibrating poker units and mechanical trowels. Other pieces of plant have been included in this study and are accounted for in the quantitative analysis of costs presented in the next section. Often contractors hire the above plant, incurring hire costs that are usually charged on a fixed daily basis.
  - Plant operating costs. These costs are a function of pour duration, fuel prices, plant running during idle periods and repair costs.
  - Labour costs. RMC is placed by labour teams, often consisting of a ganger (or foreman), plant operator, placing personnel, general labourers and masonry specialists (for finishing) and they are usually paid at an hourly rate.
  - Truck mixer surcharge. The RMC supplier often has a tolerance for the length of time that a truck mixer may remain on a project site. This tolerance is roughly 5 minutes per cubic metre. For example, a truck mixer carrying a load of 8m^3 of RMC would be permitted a time of 40 minutes on a site. Beyond this tolerance, the supplier applies a surcharge of £2/ minute over the tolerance.
  - Transport of material. This is a cost absorbed in the price of the material and is not directly applicable to the contractor, but could be useful in identifying ways to reduce the contractor's costs.

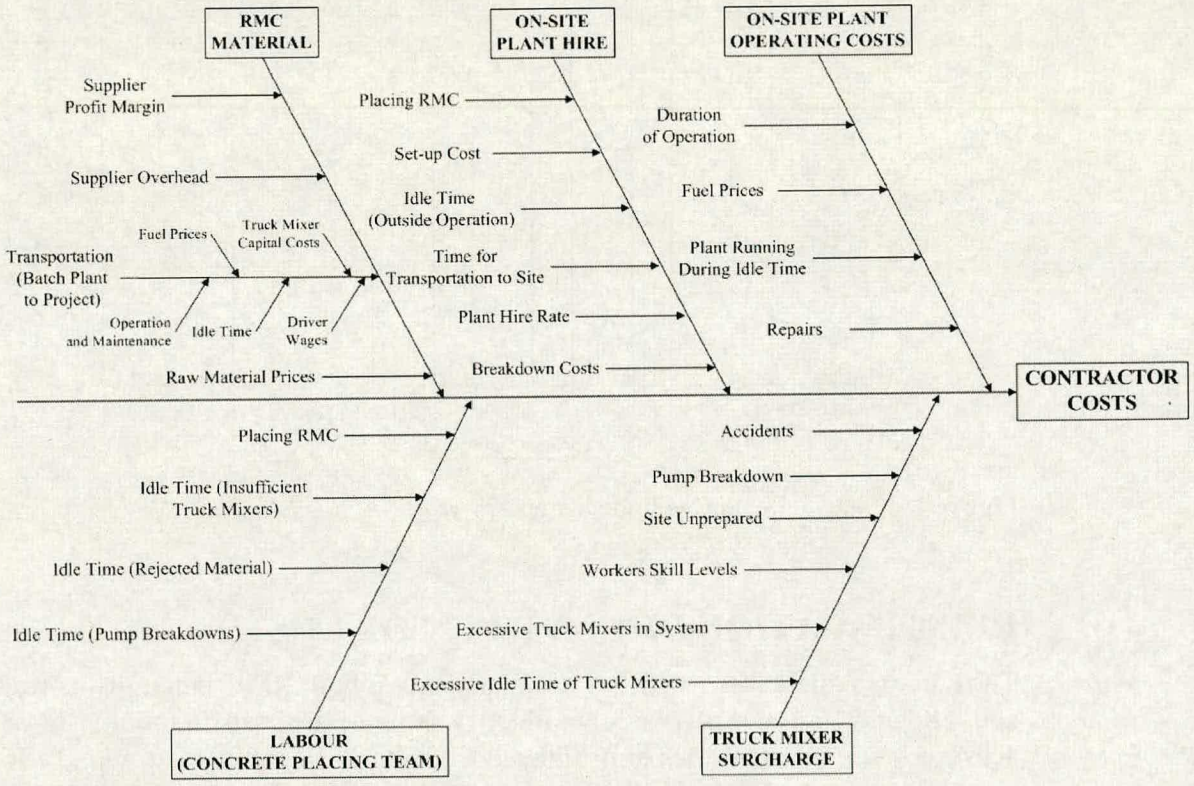


Figure 2: Ishikawa diagram of contractor costs in RMC system

QUANTIFICATION OF CONTRACTOR COSTS

The contractor costs in the RMC system identified through discussion with practitioners and presented in figure 2 were quantified for each of the 232 RMC pours to determine the largest costs in the system. These costs were calculated using the following formulae:

$$RMC\ Material\ Costs\ (£) = Unit\ Cost\ of\ RMC\ (£/m^3) \times Pour\ Volume\ (m^3) \quad (1)$$

$$Plant\ Hire\ Costs\ (£) = Daily\ Hire\ Rate\ (£/day) \times Days\ for\ Pour \quad (2)$$

$$Plant\ Operating\ Costs\ (£) = Unit\ Operating\ Cost\ (£/hr) \times Pour\ Duration\ (hrs) \quad (3)$$

$$Labour\ Hire\ Costs\ (£) = Duration\ (hours) \times [No.\ Persons \times Unit\ Cost\ of\ Each\ Job\ (£/hr)] \quad (4)$$

$$Truck\ Mixer\ Surcharge\ (£) = \begin{cases} [\text{Truck Mixer Time at Project (minutes)} - \text{Tolerance Threshold (minutes)}] * £\ 2/\text{minute} & \text{If Truck Mixer Time at Project} > \text{Tolerance Threshold} \\ 0 & \text{If Truck Mixer Time at Project} <= \text{Tolerance Threshold} \end{cases} \quad (5)$$

where, Tolerance Threshold = 5 minutes per cubic metre of RMC to be placed.

$$Contractor\ Cost\ (£) = Plant\ Hire\ Costs + Plant\ Operating\ Costs + RMC\ Material\ Costs + ... + Labour\ Hire\ Costs + Truck\ Mixer\ Surcharge \quad (6)$$

The values for pour volume, duration of pour and labour force were recorded for each pour. The unit costs and hire rates were obtained from RMC suppliers and construction contractors and are hence at 2005 prices.

A total of 168 RMC material costs, plant hire costs, plant operating costs, labour hire costs and truck mixer surcharges were calculated for the small pour project data; and

68 for large pours. The total contractor cost was £543,000 for small pours. The amount that each of the above costs contributed to this total are shown in figure 3. In summary, 73% of the contractor cost was to pay for RMC material cost. 25% was due to plant and labour costs, while the remaining 2% was due to truck mixer surcharges. It is of note that a similar breakdown of the contractor costs in large pours was recorded.

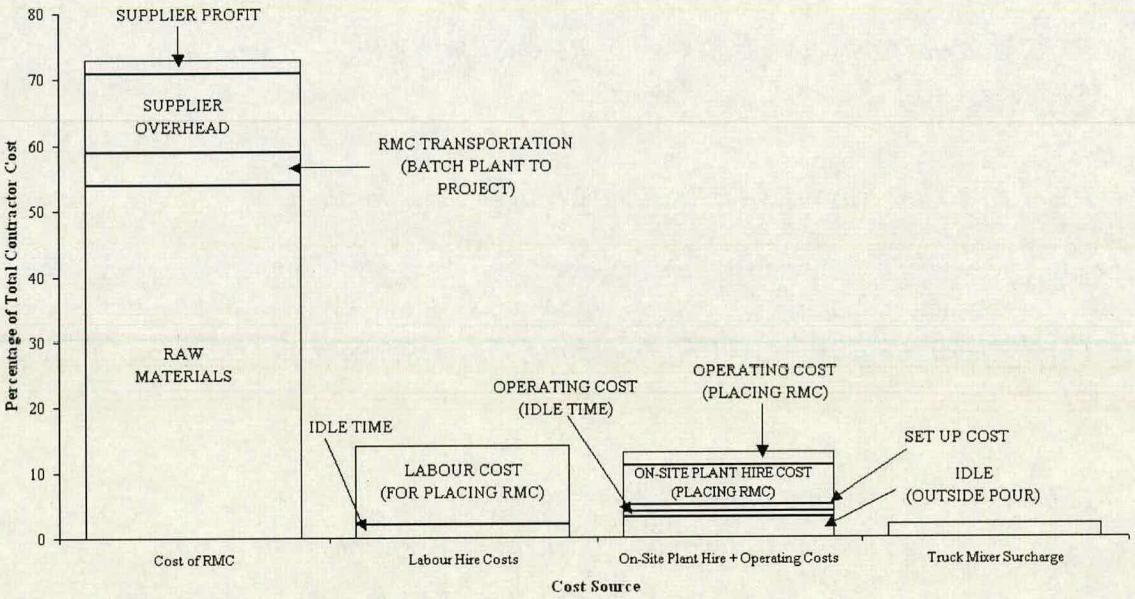


Figure 3: Contractor costs and contributors to cost in the observed small pours

COMPARISON WITH EXISTING COST INFORMATION

RS Means building construction cost data provide estimates of the breakdown of contractor costs in performing RMC operations (RS Means, 1998), and these estimates for small pours are shown in table 1. The cost breakdown from this study is similar to that presented in RS Means, which helps to provide some indication of the generic applicability of this study. However there are differences of note:

- RS Means do not recognise the impact of truck mixer surcharge on the contractor costs; a cost that was found to be significant in this study. This cost may be absorbed in the RMC material cost.
- The RS Means edition was from 1998 and is based in USA; the quantitative analysis presented here is based upon 2005 UK prices. Since 1998 the cost of RMC raw materials globally has increased (DTi, 2004). In UK this has particularly the case due to the introduction of the UK Government's Aggregates Levy in 2002 (DEFRA, 2005). Additionally, the price of crude oil has increased significantly from 1998 to 2005, resulting in increased on-site plant operating costs.

Table 1: Costs breakdown (as % of total) comparison with RS Means (after RS Means, 1998)

Contractor Costs	Estimate Source	
	Quantitative Analysis	RS Means
RMC Material	73	72
Labour Hire	13	19
On-Site Plant Hire and Operation	12	9
Truck Mixer Surcharge	2	0

CONTRIBUTORS OF CONTRACTOR COSTS IN RMC SYSTEM

The contributors to costs in the RMC system were identified through observation on project sites and quantified using the recorded data. These are presented and quantified in figure 3. The contributors to each cost in the system are discussed below.

Cost of RMC material

The contributors to RMC materials cost are as follows:

- Raw materials. Cement, fine aggregates (sand), coarse aggregates (e.g. gravel) and water are all base materials of RMC. Also, some mixes of RMC require either mineral or chemical admixtures to be added to the mix. Each of these raw materials is governed by the market price. The cost estimate of each of raw material was provided by a RMC batch plant. Raw materials account for 54% of the total contractor cost in the studied small pours.
- Transportation. When the RMC material is batched it is transported to the project in truck mixers. There are a number of factors that affect material transportation such as fuel prices, truck mixer capital, operation and maintenance costs, periods of idle time, and the driver's wages. An estimate of the unit cost of transportation was obtained from a RMC batch plant and this was used in the quantitative analysis undertaken in this research. Transportation was estimated to account for 5% of the total contractor costs (small pours).
- Supplier overhead. A number of factors contribute to the RMC supplier's overheads, such as the batch plant capital, operating and maintenance costs, labour wages, business rates, utility charges. These have been estimated to contribute to 12% of the total contractor cost in the studied small pours.
- Supplier profit. The profit margin of a RMC supplier is commonly a small percentage of the total costs. A survey of the RMC batch plants visited in this project revealed that on average the supplier's profit equates to 2% of the total contractor cost.

Labour hire cost

Labour hire costs can be broken into two groups of factors: those where labour are working on the placement of RMC (value-adding); and those where labour are standing idle for a number of reasons. For small pours, the former account for 11% of the total contractor cost, while the latter for 2% of this total cost. Three reasons for labour standing idle were identified in the observation of RMC pours: there were an insufficient number of truck mixers being used to supply the RMC resulting in gaps in the supply chain and periods were there was no RMC for the labour to place; when a load of RMC is rejected, most commonly due to failing a slump test, a gap in the supply of RMC forms as the next truck mixer will not arrive immediately – thus the

labour stands idle; when the pump suffered a breakdown the labour were forced to stand idle at a cost to the contractor. The data was analysed to quantify the costs associated with these three reasons for idleness. Although idleness due to rejected loads and pump breakdowns was observed, these events occurred infrequently and thus the costs (over all the project data) associated with this idleness were negligible. It was found that 2% of the total contractor cost could be attributed to using an insufficient number of truck mixers to delivery the RMC.

Plant hire and operating costs
The contributors to plant hire and operating costs are as follows:

- Idle plant (outside pour). Plant is usually hired for a specified period of time from a specialist supplier at a fixed rate per day. Assuming that a standard working day lasts for 12 hours, plant remains on site, but idle often for significant periods of this 12 hours, at a cost to the contractor. This cost is most invisible, because of the flat-rate hire charge. It has been estimated that this causes 3% of the total contractor cost in small pours.
- Set up costs. As with any equipment there is an associated set up cost, which incorporates a 'warming up' period. It has been estimated by contractors that setting up plant could contribute 1% of the total contractor cost in a project.
- Idle plant (operating cost). When there is a gap in the RMC supply the plant is rarely switched off, thus the contractor incurs plant operating costs for this idle period. This quantity has been calculated to be worth 1% of the total contractor cost.
- Plant hire (value-adding). This is the hire cost that can be attributed to the plant being used to place RMC and thus could be called a value-adding cost. This has been estimated to equate to 5% of the total contractor cost.
- Plant operating cost (value adding). Similarly to the previous cause of cost, this is the cost (operating in this instance) that can be directly related with the placing of RMC. This cost is worth 2% of the total contractor cost.

Truck mixer surcharge
The contributor to truck mixer surcharge is the excessive idleness of the truck mixers on the project site. The RMC supplier stipulates a tolerance for the length of time a truck mixer can remain on site and to exceed this tolerance would incur the charge. There are a number of reasons for this tolerance to be exceeded, including: pump breakdown; pump operator skill level; truck mixer driver skill level; accidents on the project site; poor site access routes; site unprepared in terms of labour availability, reinforcement bars in place or formwork in place; excessive number of truck mixers being used in the system resulting in long queues of truck mixers forming on the project site, extending the time each one is present on the site.

AREAS OF FOCUS FOR COST REDUCTION STRATEGIES

A number of costs and the contributors to cost could be focused upon to achieve a reduction in the total cost of placing RMC for a contractor. These could include:

- Cost of RMC material. This is the largest single cost for the contractor in the RMC system, representing 73% and 86% of total cost in small pours and large pours, respectively. The combined value (for both pour sizes) was £973, 700 over the studied projects. From the analysis three of the causes of this cost indicate that the cost of RMC may be reduced. These are, transportation

(5% total cost), supplier's overhead (12% total cost) and supplier's profit (2% total cost). If the supply chain were vertically integrated – the current external supplier were removed – it may be possible that some of these costs could be removed for the contractor. If the cost of RMC was reduced by only 1%, this would save the contractor £9, 800 – making any venture to achieve this meagre saving advantageous.

- Cost of hiring plant. A significant amount of cost (3% of total) is incurred for plant sitting idle on the project site outside of working hours, as operations rarely last for a full day. There is usually a sensible reason why pours cannot be performed continuously and thus this issue cannot be resolved. Rather it may be worthwhile for a contractor to consider re-negotiating the plant hire contract to one that allows plant to be hired on an hourly rate, allowing the equipment to be returned in some instances. The saving made by this strategy must be contrasted with the increase in costs due to additional set-up times (currently at 1% of total contractor cost). Another alternative may be for a contractor to purchase/lease the plant, given suitable market conditions (e.g. future demand for their services).
- Cost of plant and labour standing idle and the truck mixer surcharge. A balance should be achieved between using too few and too many truck mixers to delivery RMC in the system. Too few resulting in plant and labour waiting for RMC deliveries (3% total cost); and too many resulting in truck mixers queueing on site and the contractor incurring a surcharge for the truck mixers (2% total cost). A form of production control that achieves a balance in the number of truck mixers utilised to delivery RMC in a pour should be considered.

To achieve a minimum cost of RMC placement for a contractor a systemic solution to reduce costs should be applied. Such a solution would entail reducing all of the above costs as much as possible, with an obvious focus being placed upon the most significant source of cost. A possible solution is to follow the trend of aggregate producers and vertically integrate the resource (materials, plant, etc.) supply chain (Hill and Moore, 2004). Such integration is discussed further in Graham et al. (2005).

SUMMARY AND FURTHER CONCLUSIONS

This paper has presented the results of an investigation into the actual costs that a contractor incurs during RMC placement. It has been undertaken on the basis that such operations are almost always stochastic in nature, which makes planning and managing of them particularly difficult.

A further problem with the stochastic nature is that it can lead to difficulties in identifying and estimating the actual costs of a pour before it takes place. Perhaps because of this difficulty, previous studies have not attempted to fully identify all the costs of the concreting system but this paper proposed that this must be done if contractors' costs are to be reduced and allow the most significant and sensitive costs to be reduced.

Based on observations of actual concreting pours, which took place in Scotland, UK, this paper has presented the results into an investigation of where these costs lie in RMC placement. Initially, general cost areas were identified through discussion with

professional practitioners: both contractors and suppliers. While the supplier costs are not specifically part of this study, it is interesting to consider where these lie. Next, actual costs sources were itemised and it is proposed that that there are five separate headings: Cost of plant idle after a pour is complete; labour cost through insufficient truck mixers; labour costs through rejected loads; labour cost through pump breakdown; and cost of placing RMC.

After analysis of the 232 observed pours it was concluded that, not surprisingly, the cost of actually placing the RMC is the principle cost for the contractor. As this is the basic cost of all resources for the duration of the actual pour, and that no idle times are considered, it is noted that cost reduction in this area would require fundamental re-management of the RMC operations such as reconfiguration of the resource supply chain. The most insignificant costs were through rejected loads and pump breakdown and thus it is proposed that these are not considered areas for improvement.

Additional contractors costs were: labour idle time caused by insufficient truck mixers and plant idle time. The second of these is difficult to consider: it was previously discussed that few suppliers would be willing to reduce the minimum hire charge period (of one day) as it would transfer the risk of under-utilisation to themselves. Reduction of this cost may be achieved through the procurement of the plant; although the under utilisation will still remain its cost will be diminished because the hire rate is likely to be greater than a repayment of a capital investment.

This leaves cost as a result of insufficient truck mixers and it has been concluded that this may be reduced by closer management and control of these resources. Contractors must work closely with suppliers to achieve this aim.

A final conclusion is that a method of reducing all of the above costs should be sought as this will result in the cost of concrete placement for contractors moving toward an actual minimum.

REFERENCES

Anson, M., Tang, S. L. and Ying, K. C. (2002). Measurement of the performance of ready mixed concreting resources as data for system simulation. *Construction Management and Economics*, 20 (3), 237 – 250.

Camillo, J. (1996). Controlling the flow of truck mixers. *Concrete J.*, 14(4), 250 – 256.

DEFRA (2005). *Recycling and waste: Aggregates Levy Sustainability Fund in England*. http://defra.gov.uk/environment/waste. (Accessed: 09/05/05).

DTi (2004). *Construction Statistics Annual 2004*. HMSO, London.

Dunlop, P. and Smith, S. D. (2003). Estimating key characteristics of the concrete delivery and placement process using linear regression analysis. *Civ. Eng. and Env. Syst.*, 20 (4), 273 – 290.

Eurostat (2005). *DS-008451-PRODCOM-ANNUAL-2663100-Ready mixed concrete.* http://epp.eurostat.cec.eu.int. (Accessed: 04/05/05).

Graham, D., Tommelein, I. D., and Smith, S. D. (2005). Cost reduction through vertical integration of the in-situ concrete supply chain. 21st *Annual ARCOM Conference*, 7-9 September 2005, University of Salford, London. Under review.

Halpin, D. W. and Woodhead, R. W. (1976). *Design of construction and process operations. 1st edition.* John Wiley and Sons, Inc., New York, NY.

Hill, W. and Moore, B. (2004). *Business models in a consolidating ready mixed concrete industry. In: Concrete in Focus.* NRMCA, Silver Spring, MD.

RS Means (1998). *Building construction cost data. 56th edition.* RS Means Co., Inc., Kingston, MA.

Smith, S. D. (1998). Concrete placing analysis using discrete event simulation. *ICE Structures and Buildings*, 128 (4), 351 – 358.

Zayed, T. M. and Halpin, D. W. (2001). Simulation of concrete batch plant production. *J. Construction Engineering and Management, ASCE*, 127 (2), 132 – 141.

COST REDUCTION THROUGH VERTICAL INTEGRATION OF THE IN-SITU CONCRETE SUPPLY CHAIN

Darren Graham[1], Iris D. Tommelein[3] and Simon D. Smith[2]

[1, 2] *School of Engineering and Electronics, The University of Edinburgh, Edinburgh, EH9 3JN, UK*
[3] *Department of Civil and Environmental Engineering, The University of California at Berkeley, Berkeley, 94720, USA*

Concrete is an important, widely used material in modern-day construction. The in-situ placement of concrete is performed at a significant cost for contractors. This study aimed to outline a method(s) to reduce this cost, by focusing on the reduction of the most significant costs in concreting system: material, plant hire, and resource idleness.

An investigation was undertaken to determine if cost reduction could be achieved through a reconfiguration of the resource supply chain, in the form of vertical integration. In real terms this means that the contractor takes on the role of concrete manufacturer and on-site plant owner – removing the ready-mixed batch plant and on-site plant supplier from the existing supply chain. This would give the contractor more control over the key costs identified in the concreting system.

To permit the vertical integration of the resources supply chain a reliable alternative to manufacturing concrete in an external batch plant was required. The financial viability of using a volumetric mobile mixer as an alternative source of concrete was assessed. This type of mixer manufactures concrete at the required location on a project with a high production rate and quality of concrete.

The results of a cost/benefit analysis show that the reconfiguration of the resource supply chain can produce the desired system-wide reduction in contractor costs in concrete placement.

Keywords: cost, materials management, production management, ready-mixed concrete.

INTRODUCTION

The cost of concrete placement is significant for contractors and ways of reducing this cost should be examined. To this end, an investigation was undertaken to define the actual costs that a contractor incurs in placing concrete, based upon data from real construction projects where ready-mixed concrete (RMC) was being supplied by an external batching plant. Graham et al. (2005) presented the results of this investigation at ARCOM 2005 in a paper entitled 'Cost of Concrete Placement for Contractors.' It was found that the main sources of contractor cost are: RMC material, plant hire, labour and plant idleness and a surcharge for truck mixer idleness (on site). It was suggested that a method of reducing all of these costs, systemically, should be sought, and further that such a solution may involve vertical integration of the resource supply chain. In real terms, this means that a contractor

[1] Tel: 0131 650 7159, e-mail: simon.smith@ed.ac.uk

would take on the role of concrete supplier for its own projects and also procure the on-site plant to avoid hire costs.

This study has investigated the vertical integration of the resource supply chain from a contractor's perspective. In this article, the existing supply chain is presented. Followed by a discussion on the alternative technologies available for manufacturing concrete. Then the proposed integration of the resource supply chain is presented, and a cost/benefit analysis of effecting this integration is detailed. Finally, some limitations of the proposed cost reduction method are discussed.

EXISTING RESOURCE SUPPLY CHAIN

The observed concrete pours were supplied with RMC manufactured and delivered by an external concrete batching plant, and the RMC was placed on-site via a hired concrete pump. This resource supply chain is illustrated in the value stream map shown in figure 1.

Special symbols are used in this study to depict the RMC supply chain. They are (Rother and Shook, 1998):

-

Value-adding processes or tasks such as mixing RMC.
- △ Work-in-progress (inventory).
- An order to withdraw a product to deplete a supermarket.
- An order to produce a product to replenish a supermarket.
- A supermarket with controlled inventory.
-

First-in, First-out release of resources from a task.
- Physical pull of materials from a supermarket. This is the amount needed at the time of withdrawal, not a predetermined fixed amount.
- – – ► Flow of product.
- ⟹ Transportation of product to customer site.
- ▪▪▪▪▪▪► Product is pushed into inventory.

The following is a description of the roles and responsibilities of the players in the existing resource supply chain shown in figure 1:

- Designers (line i) specify concrete mixes either by recipe or performance (Tommelein and Li, 1999) in construction documents, which are made available to contractors (line vi) during the bid phase of a project.
- The contractor (line vi) uses this information to select placement sizes and methods (concrete pump in the observed projects) and decides upon a ready-mixed batch plant (line ii) to supply the concrete. The contractor (line vi) then schedules advance orders with the batch plant (line ii) providing approximate quantities and delivery times required. Thus securing batch plant and delivery capacity when required (NRMCA, 2001).
- The batch plant (line ii) uses the information from the contractor's advanced order to plan its raw materials inventory from its provider (line iii).
- When the placement method is chosen, the contractor (line vi) orders plant from a plant supplier (line iv). If the plant is not in stock, then the plant supplier (line iv) will need to order the plant from the plant distributor or manufacturer (line v) to provide the contractor (line vi) with the required plant by a specified date.

- Roughly 4-7 days before batching, the contractor (line vi) must call the batch plant (line ii) to confirm the details (delivery times and quantities of RMC) of the previously agreed order.
- On the day of the pour, RMC is produced by the batch plant (line ii) and transported to the project in 6 or 8 m^3 revolving drum truck mixers, which then discharge a batch of RMC into a concrete pump, which is then placed, cured and finished by the contractor (line vi).

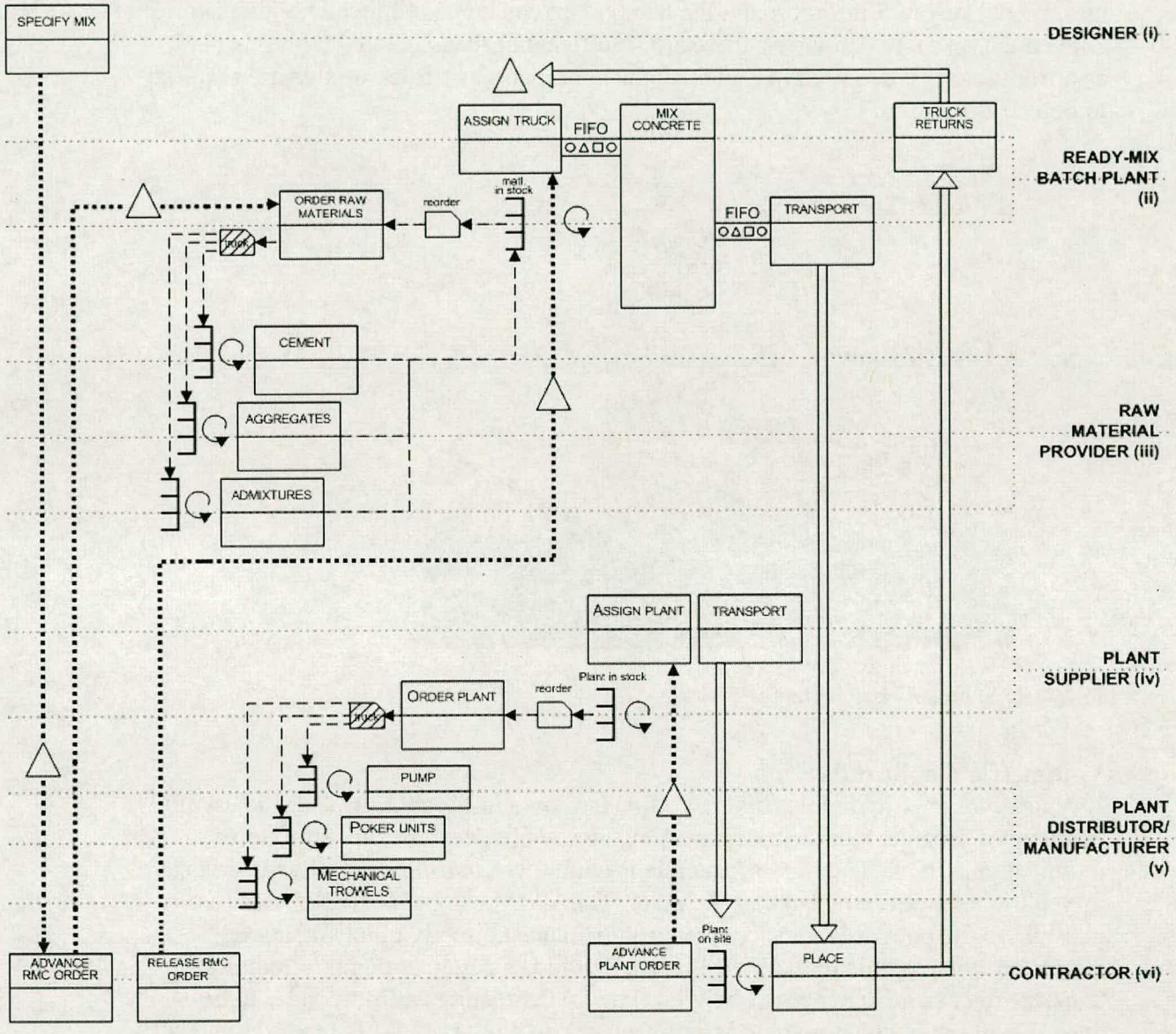


Figure 1: Value stream map of the original resource supply chain

CONCRETE MANUFACTURING TECHNOLOGIES

There are two types of mixer available for a contractor to use as an alternative to procuring RMC from a batch plant external to the project site (Sonnenberg, 1998; Ferraris, 2001): mixers that produce batches of RMC; and continuous or volumetric mixers. These two alternatives are now discussed.

Stationary batch plant

A typical stationary batch plant is shown in figure 2. It has storage bins for aggregates, a weighing bin for cement and has a cement silo attached. The required amounts of materials flow through the plant as shown in figure 2 and are placed into a truck mixer, which mixes the concrete.

The benefits of this type of plant are: that it can produce up to 68 m^3/hour; a generally acceptable rate for construction – the observed projects had a maximum placement rate of 35 m^3/hour; and the batch plant can be positioned wherever is most convenient on a site. However, there are significant costs associated with this plant, as a front loader is required to replenish the aggregates and truck mixers are required to deliver the material.

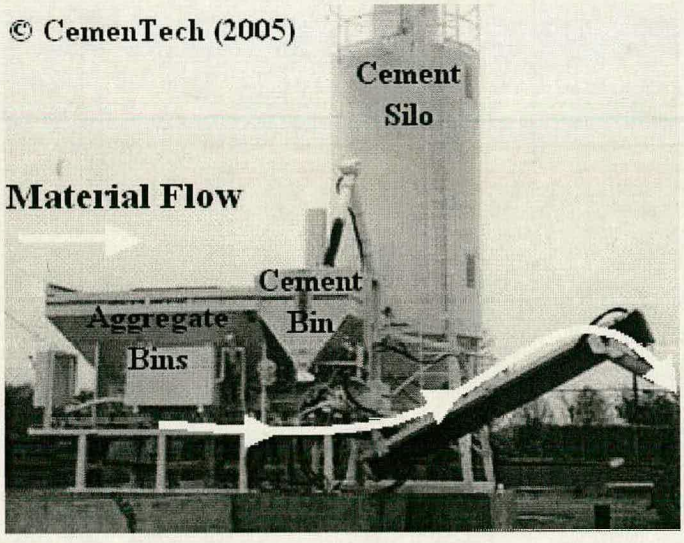


Figure 2: Stationary batch plant

Volumetric mobile mixer

A typical volumetric mobile mixer (VMM) is shown in figure 3. A VMM is usually made up of separate storage bins for aggregates, water, cement and admixtures mounted on a truck. The mix of concrete is computer-controlled and the materials are mixed in the auger at the back of the plant. The VMM can be located as close as possible to the pour area, thus concrete can be placed directly in this manner. However, it is possible that a VMM be used to feed a concrete pump to relay concrete over large distances or heights. The VMM requires a front loader to be available to maintain an inventory in the aggregate bins.A VMM is reportedly ideally suited to small pours (Ferraris, 2001). However, CemenTech, a leading manufacturer of VMM's reports that its VMM the CemenTech MCD10-200 can produce 68 m^3/hour – the same production rate as the stationary batch plant described above. Thus it is suitable for larger pours.

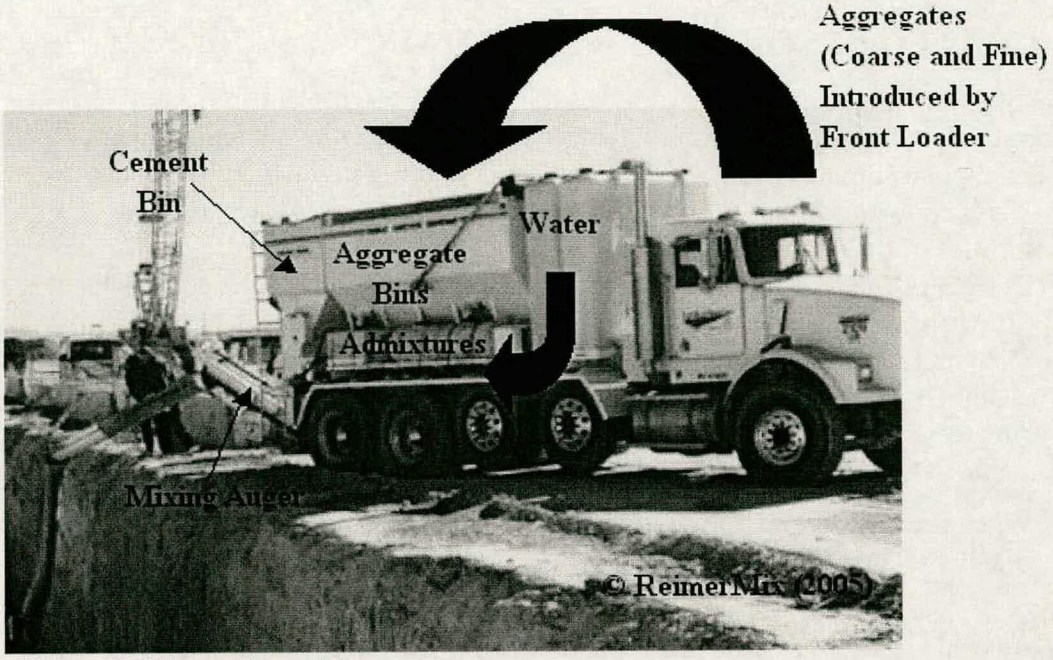


Figure 3: Volumetric mobile mixer

The VMM is not without its disadvantages. A stockpile (inventory) of aggregates would need to be kept on-hand to replenish the plant on a regular basis. Thus this method is not altogether suitable on projects were the site space is limited.

Additionally there has been some concern in the past about the quality of concrete produced by a VMM plant (Ferraris, 2001). The leading American manufacturers of VMM plant have taken notice of such criticise, and in 1999 formed the Volumetric Mixers Manufacturers Bureau (VMMB) a section of the National Ready Mixed Concrete Association (NRMCA). The VMMB aims to develop standards for volumetric mixers and agitators, and these standards have been stated in VMMB 100-01 (VMMB, 2001). These standards have been developed with existing regulations provided by:

- American Concrete Institute (ACI) in ACI 304.6R-91 (ACI, 1991) and ACI Manual of Concrete Practice, Part 1 – Materials and General Properties of Concrete (ACI, 1994);
- American Society for Testing and Materials (ASTM) in ASTM C685 produced in 1971. The latest version of this code is ASTM C685 M-01 (ASTM, 2001).
- American Association of State Highway and Transportation Officials (AASHTO) in AASHTO M-241 (AASHTO, 2004).

The VMMB report that the quality of concrete produced by VMM plant is equal, if not more consistent in quality (due to no delay in placing the material) than RMC from an external batch plant (VMMB, 2005).
When comparing the two technologies above, the need for a stationary batch plant to use truck mixers is a distinct disadvantage. This need would most likely not afford the significant reductions in contractor cost that are the aim of this study as truck mixers (even two or three) would need to be hired, leased or purchased. Thus, stationary batch plants shall not be considered further in this study as a method of implementing integration of the resource supply chain.

VMM technology shall be examined to determine if it provides a viable method of integrating the resource supply chain and reduce contractor costs in concrete placement.

INTEGRATED RESOURCE SUPPLY CHAIN

The use of a VMM by a contractor removes the external batch plant from the resource supply chain and the procurement of other plant such as a concrete pump removes the plant supplier, resulting in 3 players remaining in the integrated supply chain. This is shown in figure 4. The following is a description of the roles and responsibilities of the players in this supply chain:

- Designer (line i) provides the concrete mix specifications to the contractor (line ii) via construction documents at the bid phase.
- The contractor (line ii) selects: the placement sizes and methods; and the quantities and supplier (line iii) of raw concrete materials (sand, cement, etc.)
- An advance order is placed with the raw materials provider (line iii), based on a calculation to allow a small inventory level on the project site.
- When the contractor (line ii) is ready for placing the concrete, a front loader fills the mixer (VMM) and the concrete is placed, cured and finished.

Although the use of VMM technology and procurement of other plant has simplified the supply chain – there are now fewer players, it may be the case that the benefits of this supply chain integration may not outweigh the costs of purchasing equipment and hiring extra labour to operate the plant

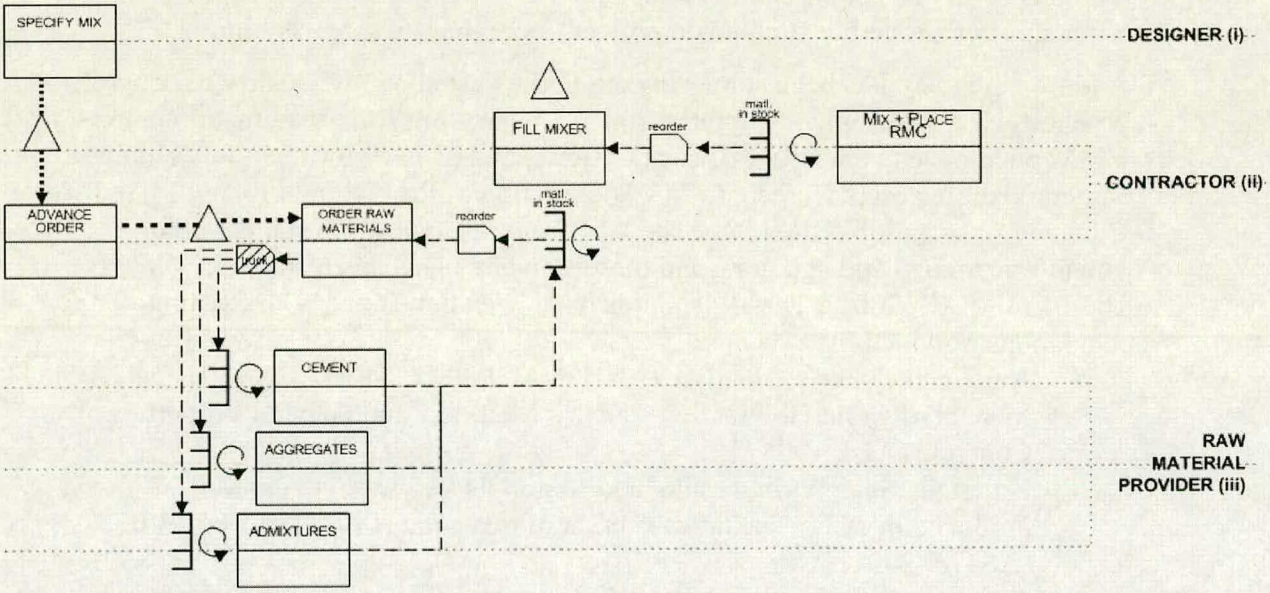


Figure 4: Proposed resource supply chain integration

COST/BENEFIT ANALYSIS OF SUPPLY CHAIN INTEGRATION

A cost/benefit analysis of integrating the resource supply chain in the manner described above has been undertaken to ascertain whether it is worthwhile. Firstly the costs were estimated, including: the capital cost of new plant; additional labour to operate the plant; and, additional plant operating costs.

Costs

To reiterate, when using a VMM the plant required is the VMM itself, a front loader and in many cases a concrete pump. Thus these are the pieces of plant considered in this cost/benefit analysis.

The cost of the new plant was obtained from estimates provided by Richie Brothers Auctioneers Inc. A VMM manufactured by CemenTech (model: MCD10-150) costs £42, 000; Caterpillar 924G Wheel (Front) Loader - £50, 000; Daewoo DCP3211 concrete pump - £70, 000.

In the cost analysis it is assumed that the plant is purchased using a business loan of £162, 000 over 15 years at a fixed rate of 7% (as quoted by HSBC UK Plc.). After 15 years the plant would be owned outright at a total cost of £332, 100 to the contractor.

The labour and plant operating costs over the 15 year period (length of loan) (table 1) are calculated based on the unit wages/operating costs presented in table 2, which take inflation into account. Additionally, it is assumed that labour will work for 10 hours/day, for 5 days/week for 4 months of a year. The costs are only for 4 months of the year as that is the length of time it took to place 6, 000 m^3 of concrete in the observed small pours.

From the above calculations the total cost of integration of the resource supply chain is £1, 060, 067 over 15 years (see table 1).

Table 1: Cost Analysis

Source of Cost	Type	Make + Model	Quantity	Cost Estimate (£)
New Plant				
	Volumetric Mobile Mixer (VMM)	MCD10-150	1	42, 000
	Front Loader	Caterpillar 924G Wheel Loader	1	50, 000
	Concrete Pump	32M	1	70, 000
				332, 100*
Additional Labour				
	VMM Operator	~	1	192, 293**
	Operator	~	1	192, 293+
				384, 586
Additional Plant Operation				
	Costs	~	1	206, 028++
	Operating Costs	~	1	137, 352+++
				343, 380
Total				**1, 060, 067**

* Inclusive of repayment of capital and interest on business loan for £162, 000 at 7% over 15 years
** Average yearly wages (Column A, Table 3) x No. hours worked/per year
+Average yearly wages (Column B, Table 3) x No. hours worked/per year
++Average yearly operating costs (Column C, Table 3) x No. hours worked/per year
+++Average yearly operating costs (Column D, Table 3) x No. hours worked/per year
No. hours worked/per year estimated as 10 hours/day, 5 days/week, 4 weeks/month and 4 months/year

Table 2: Predicted wages and operating costs

Year	Volumetric Mobile Mixer Operator Wages* £/hour	Front Loader Operator Wages* £/hour	Volumetric Mobile Mixer Operating Costs* £/hour	Front Loader Operating Costs* £/hour
2005	14.00	14.00	15.00	10.00
2006	14.27	14.27	15.29	10.19
2007	14.54	14.54	15.58	10.38
2008	14.81	14.81	15.87	10.58
2009	15.09	15.09	16.17	10.78
2010	15.38	15.38	16.48	10.99
2011	15.67	15.67	16.79	11.20
2012	15.97	15.97	17.11	11.41
2013	16.28	16.28	17.44	11.63
2014	16.58	16.58	17.77	11.85
2015	16.90	16.90	18.11	12.07
2016	17.22	17.22	18.45	12.30
2017	17.55	17.55	18.80	12.53
2018	17.88	17.88	19.16	12.77
2019	18.22	18.22	19.52	13.01

* Annual increase calculated using UK Base Rate of Inflation (1.9%)

Benefits

The benefits of resource supply chain integration have been estimated in percentage terms in table 3. These percentages were then converted into a potential reduction in unit cost (see table 3) based on an original average unit contractor cost of £53.40/ m^3 (calculated from the observed pours). The following provides a summary of the reasoning behind the estimates of the benefits (as %ages of original contractor cost):

- RMC material. There are no truck mixers involved (no transportation) giving a 5% saving; supplier overheads will be reduced significantly as there is a small number of workers directly involved, no additional administration staff, and the space required is relatively small giving a 10% saving; the supplier's profit will be absorbed by the contractor saving 2% of the original contractor's cost.
- Labour idle time. RMC is placed on demand when using a VMM, resulting in less variability in the system and the labour not standing idle waiting on concrete supply, representing a 2% saving.
- On-site plant hire and operating cost. The idle time of plant outside of the pour will still remain under the proposed integration of the supply chain. However, the loan repayment is likely to be less than the hire cost of the plant, thus resulting in a 1% saving.
- No plant hire costs – the plant is owned by the contractor giving a 5% saving.
- A VMM would not be operating during idle periods, removing the idle time operating costs present in the original RMC system (1% saving).
- Finally, as stated above there are no truck mixers in use in the system, thus there is no longer a truck mixer surcharge (2 % saving).

To summarise the costs and benefits of the proposed supply chain integration:

Total cost (over 15 year period) = £1, 060, 067 (1)

Total benefit = £14.95/m^3 (2)

Payback volume of concrete (over 15 years) = (1)/(2) = 70, 900 m^3 (3)

Annual payback volume of concrete = (3)/15 = 4, 730 m^3 (4)

Table 3: Benefit Analysis

Benefits Source	Contributor to Benefit	Potential Cost Reduction (%age Total Contractor Cost)	Value of Potential Reduction (£/m^3) (£/m^3)
RMC Material			
	Transportation of RMC (Batching Plant to Project)	5	2.67
	Supplier Overheads	10	5.34
	Supplier Profit	2	1.07
Labour Hire Cost			
	Idle Time	2	1.07
On-Site Plant Hire + Operating Costs			
	Idle Time (Outside Pour)	1	0.53
	Operating Cost (Idle Time)	1	0.53
	On-Site Plant Hire (Value-adding)	5	2.67
Truck Mixer Surcharge	~	2	1.07
Total			**14.95**

The annual payback volume is the amount of concrete that a contractor must place to "break even" on the investment. Given that the contractor placed 6, 000 m^3 in 4 months in the observed projects this appears to be an achievable target. It should be noted that plant maintenance costs are not included above.

Although the integration of the resource supply chain appears to be a promising method of reducing contractor costs in concrete placement, there are some limitations that should be highlighted. These limitations are discussed next.

LIMITS OF THE PROPOSED SUPPLY CHAIN INTEGRATION

Some of the limits of the proposed supply chain integration are:

- The plant cannot be used on multiple projects, simultaneously. This creates a plant allocation problem for a contractor.
- An area of a project site is required to store aggregates – making VMM unsuitable for use on projects with limited space.

CONCLUSIONS

This paper presented a method of reducing the system-wide costs incurred by a contractor in placing concrete. This method involved vertically integrating the resource supply chain by: the contractor appropriating the role of concrete manufacturer using a site-based volumetric mobile mixer (VMM); and the contractor procuring the plant required to place concrete (e.g. concrete pump, etc.). Thus, there would be no external RMC batch plant or plant hire company involved in the resource supply chain.

A cost/benefit analysis of this vertical integration of the supply chain was undertaken and found that the proposal was a viable method of reducing the cost of placing concrete in a systemic manner. Specifically, it was determined that a contractor would need to place 4, 700 cubic metres of concrete to make the scheme pay for itself. When compared with the studied projects, which placed 6, 000 cubic metres in four months the target volume is clearly achievable.

Some caveats to this are: a volumetric mobile mixer needs to be loaded with aggregates frequently, resulting in a requirement for a stockpile of these materials. Thus the VMM is not entirely suitable for project sites that have little available space; and the VMM can only provide concrete to one project site at a time, unlike an external RMC batch plant.

REFERENCES

AASHTO. (2004). *Report M-241. In: Standard specification for transportation materials and methods of sampling and testing.* 24th ed. AASHTO, Washington, DC.

ACI. (1991). *Guide for the use of volumetric-measuring and continuous-mixing concrete equipment: Report 304.6R-91.* ACI, Detroit, MI.

ACI. (1994). *Manual of concrete practice, part 1 – materials and general properties of concrete.* ACI, Detroit, MI.

ASTM. (2001). *Standard Specification for Concrete Made By Volumetric Batching and Continuous Mixing: Report C685/C685M-01.* ASTM, West Conshohocken, PA.

Ferraris, C. F. (2001). Concrete mixing methods and concrete mixers: state of the art. *J. Res. Natl. Inst. Stand. Technol.*, 106 (2), 391-399.

Graham, D., Smith, S. D. and Tommelein, I. D. (2005). Cost of concrete placement for contractors. 21st *Annual ARCOM Conference*, 7-9 September 2005, University of Salford, London. Under review.

NRMCA. (2001). *Industry fact sheet.* http://nrmca.org. (Accessed: 10/05/05).

Rother, M. and Shook, J. (1998). *Learning to see: value stream mapping to create value and eliminate muda.* Lean Enterprise Inst., Brookline, MA.

Sonnenberg, R. (1998). Concrete mixers and mix systems. *Conc. Precast Plant Technol.*, 64, 88-98.

Tommelein, I. D. and Li, A. E. Y. (1999). Just-in-time concrete delivery: mapping alternatives for supply chain integration. 7th *Annual IGLC Conference*, 26-28 July 1999, Berkeley, CA., 97-108.

VMMB. (2001). *Report: VMMB 100-01 Volumetric mixing standards.* VMMB, Silver Spring, MD.

VMMB. (2005). *History of the volumetric mixer.* http://vmmb.org. (Accessed: 10/05/05).