A Hybrid Fuzzy Discrete Event Simulation Framework for Analysis of Stochastic and Subjective Uncertainties in Construction Projects

by

Naimeh Sadeghi

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
CONSTRUCTION ENGINEERING AND MANAGEMENT

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
University of Alberta

ABSTRACT

Discrete event simulation (DES) has proven to be an indispensable tool for planning and analyzing construction projects. Appropriate consideration of uncertainties of the inputs of DES results in more realistic outputs. Uncertainty in general can be categorized as stochastic and subjective. Stochastic uncertainty is a system property and represents the uncertainty associated with variation of a variable. Stochastic uncertainty can be represented by a probability distribution. On the other hand, subjective uncertainty represents the lack of knowledge of the system modeller regarding the actual value of a variable. Subjective uncertainty, for example, can be a result of lack of data or linguistic expression. Subjective uncertainty is often encountered in construction simulation due to the linguistic expression, and use of expert judgment in estimating activity durations. However, traditional DES is only able to consider stochastic uncertainty using probability distributions; and cannot handle subjective uncertainty.

Fuzzy set theory provides a methodology for mathematical modelling of subjective uncertainty. Recently, fuzzy discrete event simulation (FDES) has been proposed for considering subjective uncertainty in construction simulation models. However, the fundamental differences between fuzzy numbers and probability distributions introduce new challenges to FDES frameworks. Furthermore, subjective and stochastic uncertainties may simultaneously exist in a simulation model. However, no framework is available that is able to consider both types of uncertainties in a discrete event simulation model.

Firstly, this research, proposes a methodology for considering subjective uncertainty in estimating the activity durations or productivity of construction projects. Secondly, a

FDES framework is proposed for dealing with subjective uncertainty of activity durations. The proposed framework advances the previous FDES frameworks by: (1) solving the problem of time paradox (overestimation or underestimation of the simulation time) (2) proposing a methodology for analyzing queues in FDES. Furthermore, this research proposes a novel hybrid discrete event simulation (HDES) framework that can simultaneously deal with both stochastic and subjective uncertainties. FDES framework is integrated within the proposed HDES framework for processing fuzzy uncertainty. Sampling from probability distributions are used to process stochastic uncertainty. The proposed framework is validated against analytically solved queuing examples containing both fuzzy and probabilistic uncertainty. The practicality of this framework is demonstrated using a case study of a module assembly yard. The results of this case study are compared with the results of FDES and traditional DES to demonstrate the advantages of the proposed HDES framework.

Preface

This thesis is an original work by Naimeh Sadeghi. Part of Chapter 4 of this thesis has been published as Sadeghi, N., Fayek, A. R., & Mosayebi, S. P. (2013, June). Developing a fuzzy discrete event simulation framework within a traditional simulation engine. In IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint (pp. 1102-1106). I was responsible for model development and analysis as well as the manuscript composition. Mr. Mosayebi assisted me in literature review, and Dr. A.R. Fayek was the supervisory author and was involved with concept formation and manuscript composition. Also, Appendix B of this thesis has been published as Sadeghi, N., and Fayek, A. R. (2011). A fuzzy-based approach for proactive scheduling of construction projects. Proceedings, 3rd International/9th Construction Specialty Conference, CSCE, Ottawa, Ont., June 14-17: CN-007-1-CN-007-11. I was responsible for model development and analysis as well as the manuscript composition. Dr. A.R. Fayek was the supervisory author and was involved with concept formation and manuscript composition.

ACKNOWLEDGEMENTS

I would like to thank God for surrounding me with wonderful people to perform this research project. I take this opportunity to extent my thanks to these people.

First and foremost, I thank my supervisor, Professor Aminah Robinson Fayek; without her assistance and her dedicated involvement in every step throughout the process, this thesis would have never been accomplished. I would like to thank you very much for your support and understanding over the past seven years during my M.Sc. and Ph.D. programs. I would also like to sincerely thank my supervisory committee, Dr. Yasser Mohamed and Dr. Armann Ingolfsson for their supports and invaluable comments during the different steps of this research. Thank you for your assistance in carrying out this research. Furthermore I would also like to sincerely acknowledge my external examiner, Professor Dulcy Abraham and my candidacy and doctoral committee, Dr. Mustafa Gul and Dr. Ming Lu for their thoughtful reviews and invaluable suggestions.

I also would like to take this opportunity to thank all individuals, colleagues and friends who assisted me during this research project. I extend my sincere thanks to the managers and staff of the partner company for their collaboration and support for performing this research.

Finally, I would like to express my deepest gratitude to my beloved family. My special thanks go to my parents for their never ending love and supports. Your supports, prayers, and encouragements are my reliable resources in the journey of life. I also express my appreciation to my beloved sister, Saeedeh, for her encouragements. I am grateful to my in-laws for their continual support and help. My heartfelt thanks go to my lovely son Mohammad. Your presence makes this journey much more joyful. I also extend my special acknowledgement to my loving husband, Hosein, whose companionship and support can never be appreciated enough. Thank you for your continuous love, inspiration, and patience during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 MODELLING UNCERTAINTY IN CONSTRUCTION SIMULATION

Construction processes are complex and are affected by various uncertain factors such as weather changes, breakdown of equipment, lack of skilled labour, and delayed delivery of materials. These sources of uncertainty may result in severely over budget or behind schedule construction projects. Because of the complexity of construction projects, planners often fail to consider the combined impact of various factors for managing construction projects (Ahuja and Nandakumar 1985). Thus, simulation models are extensively used in construction management to demonstrate these potential impacts. Discrete event simulation (DES) has proven to be a powerful tool for planning and analyzing construction projects (Halpin 1977). For example, DES is used for if-then analysis to consider different operational strategies and to calculate project estimates (Sadeghi and Fayek 2008, Song and AbouRizk 2006, Hajjar and AbouRizk 1996). However, accurate estimation of simulation inputs is one of the most challenging aspects of developing simulation models for construction projects. The quality of input parameters and simulation logic determines the quality of the simulation results (Maio et al. 2000). Incorrect inputs to a simulation model result in incorrect and misleading outputs (AbouRizk 2010, AbouRizk and Sawhney 1993, Chick 1999, Zhang et al. 2005). Therefore, considering uncertainties in inputs of DES is essential for having reliable simulation results.

Traditionally, probability distributions (i.e. random variables) are used to represent uncertainties for DES. Developing reliable probability distributions requires sample data of real project activities. However, in many construction projects, developing probability distributions of construction projects is not feasible due to the lack of such historical data (Zhang et al. 2005). Collecting sufficient historical data is often very expensive and time-demanding for construction projects, which make a probabilistic approach impractical when a simulation analysis needs to be carried out within a certain time limit. Additionally, collecting enough historical data may be impossible in some cases due to the uniqueness of the activities or the conditions surrounding them. Some alternative methods exist in the literature for eliciting probability distributions from expert judgement (Garthwaite et al. 2005). However, expert judgment results in an uncertainty that is due to subjectivity and linguistic expression of knowledge rather than randomness. This uncertainty cannot be adequately addressed by probability distributions (Helton 1997, Cooper et al. 1996, Dong et al. 2014, Jahani et al. 2014, Singh et al. 2010, Zadeh 2008, Zeng et al. 2014). Furthermore, the uncertainty due to linguistic expression may be also encountered when explicitly modelling factors impacting construction activity durations. The impact of different factors may be expressed by experts in an imprecise or linguistic manner.

Generally, uncertainty can be categorized as subjective and stochastic (Helton 1997). Stochastic uncertainty is a system property and represents the uncertainty associated with actual variation of a variable. On the other hand, subjective uncertainty represents the lack of knowledge of the system modeller regarding the actual value of a variable (Beer et al. 2013). Subjective uncertainty is often encountered in construction simulation due to the lack of data, linguistic expression and use of expert judgment in estimating activity durations. For example, the linguistic expression of weather as "hot" provides some information regarding the temperature, but the actual value of the temperature is not known; in this case, temperature contains subjective uncertainty. With regard to the activity durations, stochastic uncertainty corresponds to variability due to the random (stochastic) characteristics of the activities such as environment and materials variations. In contrast, subjective uncertainty results from limited knowledge due to the lack of data about the activity durations and the use of expert judgment to estimate those durations. Fuzzy set theory (Zadeh 1965) provides a methodology for mathematical modelling of subjective uncertainty (Helton 1997, Cooper et al. 1996, Dong et al. 2014, Jahani et al. 2014, Singh et al. 2010, Zadeh 2008, Zeng et al. 2014). In the rest of this dissertation fuzzy uncertainty and subjective uncertainty have been used interchangeably.

Fuzzy set theory has been used lately in many applications in the domain of civil engineering and construction management (e.g. Adeli and Hung 1994, Adeli and Sarma 2006, Hsiao et al. 2012, Jin and Doloi 2009, Lee et al. 2011, Paek et al. 1993, Rokni and Fayek 2010, Sadeghi et al. 2010, Stathopoulos et al. 2008, Yan and Ma 2013). Recently, fuzzy discrete event simulation (FDES) has been proposed for construction management as an integration of fuzzy set theory and DES (Sadeghi et al. 2013, Sadeghi and Fayek 2014, Zhang et al. 2005). FDES can deal with subjective uncertainty in construction simulation, while the traditional DES can only model stochastic uncertainty using probability distributions.

Although FDES can greatly benefit the simulation of construction management applications, the fundamental differences between fuzzy numbers and probability distributions introduce new challenges to FDES. Furthermore, subjective and stochastic uncertainty may simultaneously exist in a model (Zadeh, 2008). Fuzzy and stochastic uncertainties are proposed as complementary methods of representing uncertainty (Pedrycz and Gomide 2007). A probability distribution allows representing random behaviour (i.e. stochastic uncertainty), while fuzzy sets allow representing “partial truth, partial precision, or partial possibility” (Zadeh 2002). Fuzzy set theory along with probability theory can effectively address both subjective and stochastic uncertainty that may simultaneously exist in a system. However, a discrete event simulation framework that can handle both of these uncertainties is not currently available. Traditional DES can only handle stochastic uncertainty (represented by probability distributions) and FDES can only handle subjective uncertainty (represented by fuzzy sets).

In order to accurately capture fuzzy uncertainty along with stochastic uncertainty in practical simulation models of construction projects, a discrete event simulation framework that can handle both fuzzy and stochastic uncertainties is required. This research develops a hybrid fuzzy discrete event simulation (HFDES) framework that can handle both fuzzy and stochastic uncertainties simultaneously. However, as the first step, developing a reliable FDE simulation approach to model uncertainty due to subjectivity, vagueness, or imprecision in practical applications of construction simulation is essential.

1.2 PROBLEM STATEMENT

One of the main problems in construction simulation models that limits the use of discrete event simulation is the unavailability of accurate estimates for the probability distributions of activity durations. As discussed in the introduction, fuzzy set theory can provide an opportunity to represent the subjective uncertainty in activity durations. Although a wide variety of techniques exist in the literature for modelling fuzzy uncertainty, few researchers used fuzzy set theory to model the uncertainty of construction activity durations. Therefore the first problem is that very limited research is available for developing fuzzy sets of construction activity durations. This limitation confines the appropriate consideration of subjective uncertainty in construction activity durations in many situations.

The second problem is related to the FDES framework; FDES is different from the traditional DES in two main aspects: (1) the calculation of the event times; and (2) the selection of the next event. In FDES, fuzzy arithmetic is employed for calculating the event times. For selecting the next event, various fuzzy ranking methods have been suggested in the literature. However, there is no agreement on the selection of the most appropriate ranking method in FDES. Moreover, no objective criterion is available to determine the most appropriate methods for developing FDES of construction projects. Advancing the simulation time by the means of fuzzy ranking produces the problem of time paradox as one of the main challenges in FDES. Generally, it is expected that the simulation time only goes forward when the simulation advances, thus in any simulation state, the algebraic subtraction of previous simulation time from the current simulation time must be always positive. However, in current FDES frameworks, there is the possibility of getting negative values from this subtraction, emerging the problem of time paradox (Perrone et al. 2001).

Also, current FDES frameworks are only capable of calculating the simulation time and project completion time; methodologies for calculating other performance measures such as average queue length and waiting time have not yet been developed.

The third problem is the unavailability of a DES framework to model both fuzzy and stochastic uncertainty. There are many situations in construction simulation in which simulation inputs contain both fuzzy and stochastic uncertainties simultaneously. For example, for estimating the duration of activities in a project, we may have enough historical data for some of the more repetitive activities, but we need to use expert judgement to estimate the duration of other activities. Also, even when historical data are available for the duration of project activities, one may choose to modify the estimated value of those durations by considering some qualitative factors and expert knowledge based on the specific conditions of the project. In this case, fuzzy set theory can be employed for representing the qualitative factors and expert knowledge, while probability distributions can be developed based on historical data to model stochastic uncertainty. Therefore, it is useful to effectively combine fuzzy and stochastic uncertainties when estimating the uncertainties for a realistic simulation results. However, current event-based simulation frameworks support either stochastic uncertainty (using DES) or subjective uncertainty (using FDES) and a framework that can handle both types of uncertainties is not available.

1.3 RESEARCH OBJECTIVES

Based on the issues outlined in the above sections, this research has the following objectives:

1) To enhance approaches for estimating the durations of construction activities as inputs to simulation models:
    a. To develop new approaches for estimating uncertainty in predicting the duration or productivity of construction projects.
    b. To consider subjective uncertainty in modelling uncertainty of activity durations.
2) To enhance the state of the art of FDES for construction management:
    a. To develop a FDES engine for planning construction projects.
    b. To suggest appropriate approaches to fuzzy ranking for FDES of construction projects.
    c. To investigate approaches for calculating practical simulation outputs (e.g. waiting time, queue length) when employing the FDES for construction projects.
3) To develop an event-based simulation framework that is able to handle both fuzzy and stochastic uncertainties:
    a. To update the FDES engine to handle both fuzzy and stochastic uncertainties.
    b. To validate practical application and benefits of the proposed framework through practical examples and case studies.

1.4 EXPECTED CONTRIBUTIONS

Several contributions are presented in this thesis; some of the contributions are more relevant to researchers and classified as academic contributions. Other contributions to enhance the current practice of construction management are classified as industrial contributions.

1.4.1 Academic Contributions

The academic contributions of this research are as follows:

- Proposing a new methodology for considering uncertainties in predicting duration or productivity of construction activities.
- Developing a FDES framework for considering subjective uncertainty in construction management that eliminates the problem of time paradox from FDES.
- Proposing methodologies for analysis of queues in FDES
- Proposing a HFDES framework to consider both subjective and stochastic uncertainties in event-based construction simulation models.

1.4.2 Industrial Contributions

The industrial contributions of this research are as follows:

- Facilitating the use of simulation in construction industry when enough historical data are not available for estimating activity durations.

- Developing a framework based on fuzzy rule based systems for predicting the productivity or duration of activities of industrial construction projects.
- Developing an integrated simulation framework to consider the impact of various uncertain factors on productivity and duration of industrial construction projects.

1.5 RESEARCH METHODOLOGY

To achieve the objectives of this research, the research study in this thesis is conducted in three main stages:

1.5.1 The First Stage

Available approaches for representing subjective uncertainty of activity durations are studied. Also, a framework is proposed for developing interpretable prediction models from data to estimate the duration or productivity of construction activities. In this framework, the current state of the art is employed for developing data driven fuzzy rule-based systems. Also, a novel methodology for representing the uncertainty of the output of fuzzy rule-based systems is proposed. An example of estimating the duration of the activities of module assembly yard is also used to illustrate the practicality of the proposed approach.

1.5.2 The Second Stage

The problem of time paradox and estimating queue performance measures in FDES is investigated in details in this stage of research. A FDES framework is proposed to overcome the available shortcomings of current FDES frameworks.

The proposed FDES framework is implemented in Simphoney.Net simulation engine. The practical aspects of this framework are illustrated using building construction, tunneling, and asphalt pavement examples. Moreover, the results of FDES for finding queue performance measures are validated against analytical approaches of finding queue performance measures.

1.5.3 The Third Stage

An event-based simulation engine that can handle both fuzzy and stochastic uncertainties is developed. The proposed approach is based on sampling from stochastic uncertainty. As a result, this sampling converts HFDES to a simulation model that contains only fuzzy types of uncertainty. An FDES is utilized to perform this simulation. Approaches based on fuzzy arithmetic and fuzzy random variables for analyzing the output of HFDES for estimating the simulation time and queue performance measures are proposed. The HFDES is validated using queueing examples analyzed with available analytical solutions.

The proposed HFDES framework is applied on a case study of a module assembly yard in industrial construction. This model is developed for the partner company to estimate the productivity based on various factors affecting productivity of activities in the module yard. A fuzzy rule-based system is used to model the impact of factors on the productivity of some of the activities of module assembly yard. A data driven technique (fuzzy clustering combined with genetic algorithm) is used to develop these fuzzy rule-based systems. The fuzzy rule-based systems are integrated into the proposed HFDES to predict the overall productivity of the module assembly yard.

1.6 THESIS ORGANIZATION

A brief background, a statement of the problem, as well as the expected contributions and the methodology of this research is provided in the current chapter. The rest of this dissertation is organized as follows:

Chapter 2 provides a brief introduction to estimating uncertainty of construction activities. This chapter also reviews available methods for representing subjective uncertainty in estimating activity durations of construction projects.

In Chapter 3, the development of an interpretable data-driven productivity prediction model is proposed. This approach employs fuzzy clustering and genetic algorithm to develop an interpretable fuzzy rule-based system for predicting productivity of construction activities. Also, a new approach for representing the uncertainty of a data driven fuzzy rule-based system is presented in this chapter.

Chapter 4 provides a framework for FDES of construction projects. First, the shortcomings of available FDES frameworks are discussed. Then a new approach for FDES for calculating the event times to enhance the performance of FDES is presented. The proposed FDES approach is validated using an example of project network and a case study of tunneling.

Chapter 5 provides a methodology for analysis of queues (calculating average queue length and waiting time) in FDES. The proposed methodology is validated through mathematically solved queueing examples, and its practical aspects are illustrated using an example of an asphalt paving operation.

Chapter 6 proposes a hybrid fuzzy discrete event simulation (HFDES) framework that can consider both subjective and stochastic uncertainties. FDES framework is integrated to the proposed HFDES framework for processing fuzzy uncertainty. On the other hand, sampling from probability distributions are used to process stochastic (random) uncertainty. The proposed framework is validated through queueing examples containing both fuzzy and stochastic uncertainties. Furthermore, the practicality of the proposed framework is illustrated through the case study of module assembly yard. This chapter compares the productivity obtained with traditional DES, FDES and HFDES with the actual productivity of the module assembly yard.

Chapter 7 describes the conclusions, contributions, and limitations of this research, as well as recommendations for future research.

1.7 REFERENCES

AbouRizk, S. (2010). Role of simulation in construction engineering and management. Journal of Construction Engineering and Management, 136(10), 1140-1153.

AbouRizk, S. M., & Sawhney, A. (1993). Subjective and interactive duration estimation. Canadian journal of civil engineering, 20(3), 457-470.

Adeli, H., & Hung, S. L. (1994). Machine learning: neural networks, genetic algorithms, and fuzzy systems. John Wiley & Sons, Inc.

Adeli, H., & Sarma, K. C. (2006). Cost optimization of structures: fuzzy logic, genetic algorithms, and parallel computing. John Wiley & Sons.

Ahuja, H. N., & Nandakumar, V. (1985). Simulation model to forecast project completion time. Journal of construction engineering and management, 111(4), 325-342.

Beer, M., Ferson, S., & Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. Mechanical Systems and Signal Processing, 37(1), 4-29.

Chick, S. E. (1999, December). Steps to implement Bayesian input distribution selection. In Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future-Volume 1 (pp. 317-324). ACM.

Cooper, J. A., Ferson, S., & Ginzburg, L. (1996). Hybrid processing of stochastic and subjective uncertainty data. Risk Analysis, 16(6), 785-791.

Dong, C., Tan, Q., Huang, G. H., & Cai, Y. P. (2014). A dual-inexact fuzzy stochastic model for water resources management and non-point source pollution mitigation under multiple uncertainties. Hydrology and Earth System Sciences Discussions, 11(1), 987-1022.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. Journal of the American Statistical Association, 100(470), 680-701.

Hajjar, D., & AbouRizk, S. M. (1996). Building a special purposes simulation tool for earth moving operations. Proceedings of the 28th conference on Winter simulation (pp. 1313-1320). IEEE Computer Society.

Halpin, D. W. (1992). Planning and analysis of construction operations. John Wiley & Sons.

Helton, J. C. (1997). Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. Journal of Statistical Computation and Simulation, 57(1-4), 3-76.

Hsiao, F. Y., Wang, S. H., Wang, W. C., Wen, C. P., & Yu, W. D. (2012). Neuro-Fuzzy Cost Estimation Model Enhanced by Fast Messy Genetic Algorithms for Semiconductor Hookup Construction. Computer- Aided Civil and Infrastructure Engineering, 27(10), 764-781.

Huang, W., Ding, L., Wen, B., & Cao, B. (2009). Project scheduling problem for software development with random fuzzy activity duration times. In Advances in Neural Networks–ISNN 2009 (pp. 60-69). Springer Berlin Heidelberg.

Jahani, E., Muhanna, R. L., Shayanfar, M. A., & Barkhordari, M. A. (2014). Reliability Assessment with Fuzzy Random Variables Using Interval Monte Carlo Simulation. Computer‐ Aided Civil and Infrastructure Engineering, 29(3), 208-220.

Jin, X. H., & Doloi, H. (2009). Modeling risk allocation in privately financed infrastructure projects using fuzzy logic. Computer‐ Aided Civil and Infrastructure Engineering, 24(7), 509-524.

Lee, D. E., Lim, T. K., & Arditi, D. (2011). An expert system for auditing quality management systems in construction. Computer‐ Aided Civil and Infrastructure Engineering, 26(8), 612-631.

Li, Y., & Liu, C. (2012). Labour productivity measurement with variable returns to scale in Australia's construction industry. Architectural science review, 55(2), 110-118.

Maio, C., Schexnayder, C., Knutson, K., & Weber, S. (2000). Probability distribution functions for construction simulation. Journal of Construction Engineering and Management, 126(4), 285-292.

Paek, J. H., Lee, Y. W., & Ock, J. H. (1993). Pricing construction risk: fuzzy set application. Journal of Construction Engineering and Management, 119(4), 743-756.

Pedrycz, W., & Gomide, F. (2007). Fuzzy systems engineering: toward human-centric computing. John Wiley & Sons.

Perrone, G., Zinno, A., & La Diega, N. (2001). Fuzzy discrete event simulation: A new tool for rapid analysis of production systems under vague information. Journal of Intelligent Manufacturing, 12(3), 309-326.

Rokni, S., & Fayek, A. R. (2010). A multi-criteria optimization framework for industrial shop scheduling using fuzzy set theory. Integrated Computer-Aided Engineering, 17(3), 175-196.

Sadeghi, N. and Fayek, A.R. (Submitted), A New Approach for Calculating the Event Times in Fuzzy Discrete Event Simulation for Construction Applications, Journal of Automation in Construction.

Sadeghi, N., & Fayek, A. R. (2008). A framework for simulating industrial construction processes. Proceedings of Winter Simulation Conference 2008, Miami, FL, pp. 2396-2401, IEEE.

Sadeghi, N., Fayek, A. R., & Mosayebi, S. P. (2013). Developing a fuzzy discrete event simulation framework within a traditional simulation engine. Proceedings of

2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), Edmonton, AB, pp. 1102-1106, IEEE.

Sadeghi, N., Fayek, A. R., & Pedrycz, W. (2010). Fuzzy Monte Carlo simulation and risk assessment in construction. Computer- Aided Civil and Infrastructure Engineering, 25(4), 238-252.

Singh, A., Walker, D. D., Minsker, B. S., & Valocchi, A. J. (2010). Incorporating subjective and stochastic uncertainty in an interactive multi-objective groundwater calibration framework. Stochastic Environmental Research and Risk Assessment, 24(6), 881-898.

Song, L., & AbouRizk, S. M. (2008). Measuring and modeling labor productivity using historical data. Journal of Construction Engineering and Management, 134(10), 786-794.

Sonmez, R., & Rowings, J. E. (1998). Construction labor productivity modeling with neural networks. Journal of Construction Engineering and Management, 124(6), 498-504.

Stathopoulos, A., Dimitriou, L., & Tsekeris, T. (2008). Fuzzy modeling approach for combined forecasting of urban traffic flow. Computer- Aided Civil and Infrastructure Engineering, 23(7), 521-535.

Taghaddos, H., AbouRizk, S. M., Mohamed, Y., & Hermann, U. (2009). Integrated simulation-based scheduling for module assembly yard. Proceedings of the 2009 ASCE Construction Research Congress (CRC), Seattle, WA, pp. 1270-1279.

Yan, L., & Ma, Z. M. (2013). Conceptual design of object-oriented databases for fuzzy engineering information modeling. Integrated Computer-Aided Engineering, 20(2), 183-197.

Zadeh, L. A. (1965). Fuzzy sets, Information and Control, 8(3), 338-353.

Zadeh, L. A. (2002). Toward a perception-based theory of probabilistic reasoning, rough sets and current trends in computing, Malvern, PA, USA, pp. 46-48.

Zadeh, L. A. (2008). Is there a need for fuzzy logic? Information Sciences,178(13), 2751-2779.

Zeng, Z., Xu, J., Wu, S., & Shen, M. (2014). Antithetic Method- Based Particle Swarm Optimization for a Queuing Network Problem with Fuzzy Data in Concrete Transportation Systems. Computer- Aided Civil and Infrastructure Engineering, 29(10), 771-800.

Zhang, H., Tam, C. M., & Li, H. (2005). Modeling uncertain activity duration by fuzzy number and discrete-event simulation. European Journal of Operational Research, 164(3), 715-729.

CHAPTER 2 MODELLING UNCERTAINTIES IN CONSTRUCTION ACTIVITY DURATIONS

2.1 INTRODUCTION

Activity durations are the main uncertain inputs to simulation models of construction processes. Appropriate modelling of activity durations is essential as all subsequent calculations in construction simulation models are based on these durations (Ayyub and Haldar, 1984).

In some of the applications of construction simulation, the productivity is first estimated and the activity duration is derived from the productivity (e.g. Corona-Suárez et al. 2014, Shaheen et al. 2005, Song and AbouRizk 2008). This is because, in many construction activities, the quantity of the job, number of workers, or duration of shift greatly varies from activity to activity or project to project. As a result, the impact of these factors along with the productivity is considered to provide a better estimate for the activity duration as the simulation input. For this reason, the estimation of labour productivity and duration of construction activities are closely related and both of them are discussed in this chapter.

Two schools of thought exist in modelling productivity or duration of construction activities (AbouRizk and Sawhney 1993): (1) one uncertain variable is presented for the productivity or duration of construction activities which implicitly aggregates the impact of numerous factors; (2) the impact of some of the significant influencing factors on productivity or duration of construction activities are modelled explicitly using a prediction model. A review on each of these schools of thought is provided in the following sections.

2.2 UNCERTAIN VARIABLES FOR MODELLING PRODUCTIVITY OR DURATION OF CONSTRUCTION ACTIVITIES

In this school of thought, productivity or duration of activities are modelled as an uncertain variable in which the impacts of numerous factors are implicitly considered. Two scenarios may be considered for developing these uncertain variables: (1) enough historical data are available to develop a probability distribution for duration or productivity of construction activities; (2) enough historical data are not available and expert knowledge should be used to estimate an uncertain variable for the productivity or duration of construction activities. The first scenario is discussed in Section 2.2.1. In the second scenario, some researchers suggest using expert knowledge to develop a probability distribution for duration or productivity of activities. Other researchers question the quality of a probability distribution estimated using expert knowledge. These researchers propose the use of fuzzy set theory to model the productivity or duration of construction activities. These methods are discussed in sections 2.2.2 and 2.2.3 respectively.

2.2.1 Developing Probability Distributions from Historical Data

When enough historical data are available for the productivity or duration of a construction activity, probability theory is often employed; a probability distribution is developed based on the historical data to represent the uncertainty of the productivity or duration of that activity.

A probability distribution assigns a probability to the possible outcomes of an event. The outcome of the event referred as a random variable and contains stochastic uncertainty. A random variable is the most common method for representing uncertainty; it is defined in terms of probability theory. Given an experiment with a possible set of outcomes **Ω**, a random variable *X* is a function from the sample space **Ω** to the real line R (Bertsekas and Tsitsiklis 2002, Liu 2009). Pr(X=x) is the probability that a random variable X is equal to a specific value x. A random variable is continuous if the probability space **Ω** is not countable and is expressed in terms of intervals. A Probability Density Function (PDF) denoted as f_X is defined for a continuous random variable (Figure 2.1). Equation 2.1 is used to calculate the probability that the continuous random variable *X* falls within an interval (a, b) (Bertsekas and Tsitsiklis 2002).

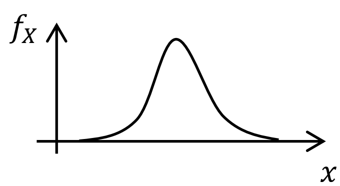


Figure 2.1. An Example of PDF of a continuous random variable

$$\Pr(a < X < b) = \int_a^b f_X(x)\, dx \tag{2.1}$$

Uncertain construction activity durations are most commonly defined using continuous random variables. When enough historical data are available, probabilistic approaches are used to estimate PDF of construction activity durations. Different approaches are available for estimating PDF from data such as Parzen windows (Parzen 1962). The most basic approach for developing PDF is based on histograms (Freedman and Diaconis 1981, Scott 1979, Shimazaki and Shinomoto 2007).

The previously discussed approaches for estimating PDF are considered non-parametric since an underlying PDF is not assumed for the activity duration. In most construction simulation applications, however, the underlying probability distribution function is assumed and the parameters of the distribution functions are estimated from data. The type of the underlying PDF of the productivity or duration of construction activities is often not known. Thus, one has to assume a type of PDF. It is recommended that the type of the underlying PDF be selected from a flexible family of PDFs that are able to represent a wide variety of shapes. Furthermore, the PDF used for the activity duration for construction simulation should be limited between two positive durations (MacCrimmon and Rayvec 1964). Based on these mentioned characteristics, as well as experiments, beta PDF have been proposed as one of the most suitable distributions for representing construction activity durations (AbouRizk and Halping 1992, Fente et al. 1999). The beta PDF is defined with four parameters: parameters a and b define the shapes of the beta distribution, while parameters minimum (min) and maximum (max) define the bounds for the distribution.

Above discussed methods required historical data for estimating the PDF of productivity or duration of construction activities. In the next section, the approaches for estimating activity durations when enough historical data for developing PDF is not available are discussed.

2.2.2 *Developing Probability Distributions from Expert Knowledge*

Expert knowledge is often used for estimating productivity or duration of construction activities when enough historical data are not available. Some experts proposed developing PDF of productivity or duration based on expert knowledge. Different methods exist in the literature for eliciting probability distributions based on expert judgment (Garthwaite et al. 2005). AbouRizk and Halpin (1992) developed a software tool called VIBES (Visual Interactive Beta Estimation System) for eliciting beta PDF of construction activity durations based on expert judgement. VIBES develop the probability distribution based on the minimum and maximum possible value of the activity duration and two other characteristics that can be any of the following: 1) mean and standard deviation, 2) mean and a selected percentile, 3) mode and a selected percentile, 4) Two selected percentile.

Although different techniques are available to facilitate the elicitation of PDF from expert judgment, the use of expert judgment often results in errors and inaccuracy in the estimated parameters of PDFs (Fente et al. 1999). Peterson and Miller (1964) specially argued that estimation of mean and standard divisions is very difficult for experts when the PDF of productivities or durations of construction activities are skewed. Many errors in the output of construction simulation models are due to assigning wrong values to the parameters of the input probability distributions. A further error occurs in the simulation output if the type of the distribution is also mistakenly assumed (e.g. assuming beta distribution when another type of distribution would have been appropriate). However, the error due to the wrong form of distribution is small comparing to the error due to the inaccurate parameters of input PDFs (Weiler 1965). The lack of accuracy of the PDFs of the activity durations that are estimated subjectively is still one of the most important limiting factors for using discrete event simulation as a planning tool in construction industry (Zhang et al. 2005).

As discussed in this section, inaccuracy in the estimated parameters of PDFs or/and wrong assumptions regarding the type of PDFs are encountered when expert knowledge are used to estimate PDFs. Thus, some researchers question the suitability of probability theory for modelling productivity or duration of construction activities when using expert knowledge. Fuzzy set theory has been proposed as a better alternative for estimating activity durations in these situations. A brief background on estimating construction productivity or duration based on fuzzy set theory is provided in the following section.

2.2.3 Developing Fuzzy Numbers from Expert Knowledge

Many researchers question the use of expert knowledge for accurate estimation of the parameters of PDF (Cooper et al. 1996, Dong et al. 2014, Helton 1997, Jahani et al. 2014, Singh et al. 2010, Zadeh 2008, Zeng et al. 2014). In this point of view, the use of expert knowledge introduces subjective uncertainty to the estimated values. Subjective uncertainty is different from stochastic uncertainty. This is because subjective uncertainty is originated from lack of knowledge. On the other hand, stochastic uncertainty is originated from randomness.

Fuzzy set theory provides a methodology for modelling subjective uncertainty. Therefore, fuzzy set theory has been proposed for representing the uncertainty of productivity or duration of construction activities when those values are derived based on expert knowledge. A fuzzy set A is defined on the universal set U by assigning a membership degree between 0 and 1 to each member of U. The membership degree indicates the degree that the members are compatible with the properties of the fuzzy set (Zadeh 1965). The membership function of a fuzzy set A is denoted as μ_A. for any $x \in U$, $\mu_A(x)$ represents the possibility (degree of membership) of x in A (Figure 2.2).

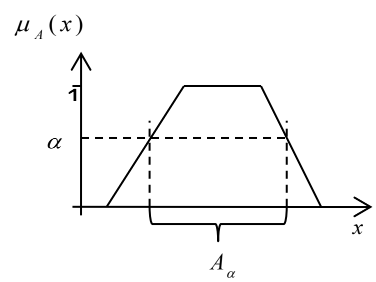


Figure 2.2 A fuzzy membership function and its alpha-cut

The alpha-cut of a fuzzy set A at the level of $\alpha \in (0,1]$ is a set A_α, whose members have a membership degree greater than α (Figure 2.2). The support of a fuzzy set is a set whose members have a membership degree greater than 0. Each fuzzy set can be reconstructed from its alpha-cuts according to the representation theorem using Equation 2.2. In this Equation, $\mu_{(B\alpha)}(x)$ represents the membership function of the interval B_α. Therefore, $\mu_{(B\alpha)}(x)$ is 1 if $x \in B_\alpha$ and is 0 if $x \notin B_\alpha$.

$$\mu_B(x) = \sup_{\alpha \in (0,1]} \alpha \mu_{B_\alpha}(x) \qquad (2.2)$$

A fuzzy number is a fuzzy set if its membership function is defined on real numbers; is piecewise, continuous, and convex; and has at least one element with full membership (maximum membership degree of one). Both fuzzy numbers and intervals allow us to represent imprecise quantities and to represent our perception of reality. However, fuzzy numbers can be viewed as a generalized form of intervals as they allow the boundaries of interval to be defined imprecisely by the means of membership degrees (Pedryz and Gomide 2007). In a similar point of view, fuzzy sets are viewed as a nested set of intervals with different membership degrees, where each interval represents an alpha-cut of the fuzzy set (Beer et al. 2013).

Fuzzy numbers have been used to represent the durations of construction activities when subjective knowledge of expert is used in estimating activity durations. An uncertain activity duration, D, can be described by a fuzzy number with membership function $\mu_D(x)$. This membership function represents the degree of possibility (membership) for the activity to have a duration equal to x units of time (e.g., days, hours). The activity duration is directly estimated by one expert or a number of experts using one of the various methods exist in the literature for developing fuzzy numbers based on expert judgment (Pedrycz and Gomide 2007). A simple method is suggested by Zhang et al. (2005) for developing activity durations, in which the activity duration is described as "most likely between $D_{m1}$and D_{m2}, but definitely not less than D_1 and not greater than D_2". A trapezoidal fuzzy number denoted as trap(D_1, D_{m1}, D_{m2}, D_2) can be developed for the activity duration based on Equation 2.3 (Figure 2.3-a).

$$\mu_D(x) = \begin{cases} \frac{x-D_1}{D_{m1}-D_1}, D_1 < x < D_{m1} \; and \; D_1 \neq D_{m1} \\ 1 \qquad , D_{m1} < x < D_{m2} \\ \frac{x-D_2}{D_{m2}-D_2}, D_{m2} < x < D_2 \; and \; D_2 \neq D_{m2} \\ \qquad 0 \qquad , \; otherwise \end{cases} \tag{2.3}$$

Developing a triangular fuzzy number for the activity duration is also proposed by Zhang et al. (2005). A triangular fuzzy number is a special case of a trapezoidal fuzzy number where $D_{m1} = D_{m2}$. A triangular number is denoted as tri(D_l, D_m, D_2) (Figure 2.3-b). Although many types of fuzzy numbers have been used to describe uncertainties, triangular and trapezoidal fuzzy numbers are the most common forms of fuzzy numbers. This is because the parameters in triangular and trapezoidal fuzzy numbers can be easily specified using linguistic terms such as “most likely” and “minimum”. Also, the arithmetic of triangular and trapezoidal fuzzy numbers is computationally more efficient compared to other shapes of fuzzy numbers. Furthermore, it has been shown that the simplicity of triangular (and trapezoidal) fuzzy numbers does not preclude their efficiency in representing subjective uncertainty (Pedrycz 1994). In fuzzy sets the impact of the type of membership function less compared with the type of probability distribution. This is because in fuzzy set theory the uncertainty is propagated in a more conservative manner compared with probability theory.

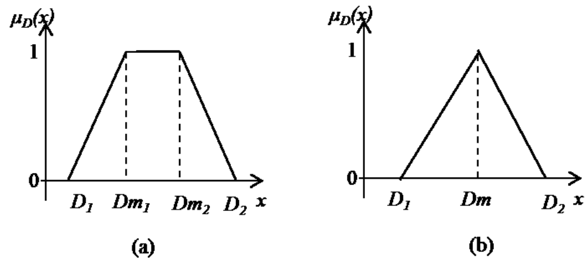


Figure 2.3 Examples of fuzzy activity durations: (a) a trapezoidal fuzzy number; (b) a triangular fuzzy number

The discussed approach of estimating four (or three) parameters for developing trapezoidal (or triangular) fuzzy numbers is one of the simplest approaches for estimating activity durations from expert knowledge. Various other methods can be employed to estimate fuzzy numbers based on expert knowledge. Some of the most common methods are as following (Pedrycz and Gomide 2007):

- *Ideal Prototype method*: In this method, the expert(s) directly estimate the membership degree of elements $x_1, x_2, \ldots x_n$ in D.
- *Horizontal Method:* For elements $x_1, x_2, \ldots x_n$ *in* D and assuming that n number of experts are available, each expert is asked if it is possible that the activity duration be x_i. The membership degree of x_i in D is estimated based on the ratio of experts who replied yes.
- *Vertical Method*: In this method, the expert is asked to identify intervals of values for the activity duration with a certain level of confidence associated with them. Each estimated interval will represent an alpha-cut of the fuzzy set. The representation theorem (Equation 2.2) will be used to estimate the final fuzzy number of the activity duration.

- *Pairwise comparison:* The direct estimates of the membership degrees are replaced by pairwise comparisons in this method. Thus, the expert is asked to compare elements $x_1, x_2, \ldots x_n$ in pairs according to their relative weights. The membership function is then estimated based on the provided weights.

As discussed in the current and previous sections, both fuzzy and probabilistic methods have been used to represent the uncertainty of productivity and duration of construction activities depending on the availability of data. One of the main reasons of uncertainty of the productivity and durations of construction activities is the numerous factors impacting those durations. Many researchers propose to use prediction models that explicitly consider the factors impacting productivity and durations of construction activities to reduce the uncertainty and provide a more accurate estimate. These approaches are discussed in the following section.

2.3 PREDICTION MODELS FOR ESTIMATING PRODUCTIVITY OR DURATION OF CONSTRUCTION ACTIVITIES

The base estimate for productivity or duration of construction activities is often subject to modification according to the specific characteristics of a project or activity, or their surrounding conditions. In fact, the duration and productivity of an activity can significantly change due to the impact of many influencing factors such as weather, skill level, and complexity. As a result, for more accurate estimation of construction activity duration and productivity, many researchers suggested to explicitly model the influence of different factors. In these approaches, a prediction model is developed for the activity duration that its inputs are some of the significant influencing factors on productivity or duration of construction activities. The output of this prediction model is productivity or activity duration. Two scenarios can be considered when developing such prediction models: (1) enough historical data are available for the activity durations and their influencing factors. In this scenario, machine-learning techniques can be used to train a prediction model; (2) enough historical data are not available for the activity durations and their impacting factors. In this scenario, expert knowledge should be employed to develop a prediction model. Following sections provides a brief review of the available approaches in each scenario.

2.3.1 Data-Driven Productivity or Duration Prediction Models

Machine learning is generally referred to as the construction of models that can learn from data. The developed model can be used for making predictions or decisions (Bishop 2006). When historical data are available for both influencing factors and productivity or duration of construction activities, machine-learning methods can be used to develop a productivity or duration prediction model.

Different machine learning methods have been used for estimating construction labour productivity. For example, Smith (1999) used linear regression to estimate earthmoving productivity. The most popular machine learning method that has been applied in predicting productivity or duration of construction activities is artificial neural network (ANN). ANN has been used for estimating construction labour productivity since 1990s (AbouRizk and Wales 1997, Karshenas and Feng 1992, Moselhi et al. 1991) and until very recently (e.g. Chang et al. 2014, Gerek et al. 2014).

ANNs are inspired from nervous systems of animals to predict one or more output(s) based on a large number of inputs. An ANN is structured as a set of neurons (or nodes) that are connected with arrows. ANN has at least three layers, one input layer, one or more hidden layer(s), and one output layer; the output of each neuron in one layer is the input to the neurons of next layer. For example, Figure 2.4 represents an ANN with three layers.

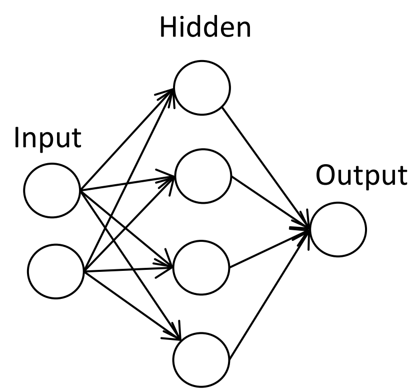


Figure 2.4 An example of an ANN network with three layers

ANN can learn from data. In learning, the structure of the network usually remains unchanged and the parameters of the nodes and arrows are updated to minimize the cost function. The commonly used cost function is the mean square error that is calculated based on the output of the network and target value.

ANNs have been proven as indispensable tools for developing accurate prediction models from data. However, one of the main disadvantages of ANN is that they are black boxes; one can "create a successful net without understanding how it worked" (Gurney 1997). Also, ANN, like any other machine learning method is dependant to large number of data to learn from. When such data are not available for estimating construction labour productivity, expert knowledge can be used to model the impact of different factors on labour productivity as discussed in the next section.

2.3.2 Developing Productivity or Duration Prediction Models Using Expert Knowledge

For developing a productivity or duration prediction model using machine-learning method, large amount of historical data should be available both for the output (productivity or duration of construction activity) and inputs (influencing factors). However, it is very common for such data not to be available for various reasons, some of which are as follows: (1) Data collection of construction activities are often time consuming; (2) Many construction projects are unique and cannot rely on historical data for estimation of the productivity or duration of their activities; (3) Construction activities are often affected by some factors that are expressed qualitatively rather than quantitatively. For example, a commonly acceptable and standard numerical value cannot be attached to weather conditions. On the other hand, weather conditions can be often described as good or poor. Other qualitative factors such as skill level of labourers and complexity of the activity are common in construction projects. Although the data regarding qualitative factors can be collected using surveys, these data are often not available for previous projects or can be expensive to collect.

In the lack of enough data, the impact of factors on the activity durations is often expressed linguistically by experts. For example, an expert may provide a statement that when the weather conditions are poor, the duration of an activity is very large. Ayyub and Haldar (1984) provide first attempts to explicitly model the impact of factors on the duration of construction activities from expert judgment using fuzzy set theory. In later developments, fuzzy rule-based systems have been employed by Fayek and Oduba (2005) and Shaheen et al. (2009) to estimate the impact of different factors on labour productivity and durations of construction activities.

Fuzzy rule-based system (also referred as fuzzy expert system) is a mathematical modelling that maps inputs to output(s). It is based on a set of rules in which fuzzy set concepts are incorporated. The antecedents of the rules in fuzzy rule-based systems are expressed linguistically. The consequent(s) of the rules may be expressed linguistically (Mamdani-type fuzzy rule-based systems) or with a mathematical function (Takagi-Sugeno-type fuzzy rule-based system) (Mamdani 1977, Takagi and Sugeno 1985).

Mamdani-type fuzzy rule-based systems are more suitable when data are not available and expert knowledge is the only source for developing the fuzzy rule-based system. This is because; the rule-based can be expressed fully in linguistic terms in Mamdani-type fuzzy rule-based systems. On the other hand, the estimation of a mathematical function that is required in the Takagi-Sugeno-type fuzzy rule-based system is often difficult for experts. A Mamdani-type fuzzy rule-based system contains four main components (Jang 1993) (Figure 2.5):

1) Knowledge base component that contains linguistically expressed if-then rules as well as a database. The database defines the membership function of the linguistic terms
2) Fuzzification component matches the crisp inputs of the model to the degrees of truth of linguist terms

3) Fuzzy inference component aggregates different rules based on their degrees of truth to estimate a fuzzy set for the output.

4) Defuzzification component transforms the fuzzy results of the model to a crisp output.

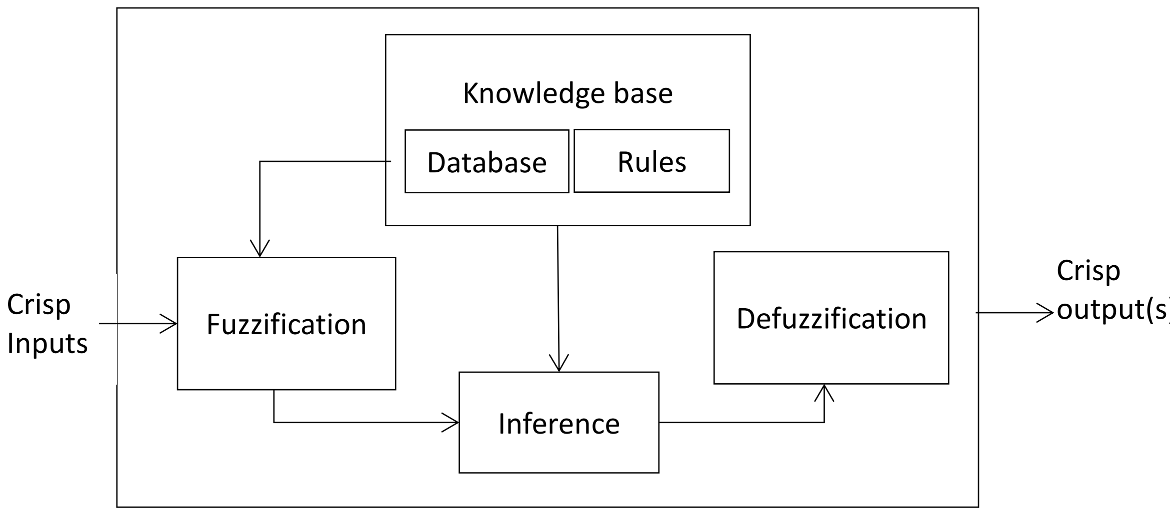


Figure 2.5 Components of a fuzzy rule-based system

For developing a fuzzy rule-based system in the lack of data, the knowledge base has to be developed based on expert knowledge: first a knowledge engineer gets data from the expert and establishes the knowledge base. Then the expert evaluates the knowledge base and gives feedbacks to the knowledge engineer. This procedure continues until a satisfactory result is obtained. As an example of the components of the knowledge base in a fuzzy rule-based system, assume the fuzzy rule-based system contains tow input factors, crew sill level and complexity, and one output, activity productivity. Productivity can be categorized to five linguist terms: very low, low, medium, high, and very high. The linguistic terms for crew skill level and complexity can be categorized to three linguist terms: low, medium, and high. The membership function of each linguistic term is developed using expert judgment. For developing these membership functions, different available methods such as horizontal method or vertical method (which are briefly discussed in Section 2.2.3) can be employed. The if-then rules in the knowledge base of a Mamdani-type fuzzy rule-based system are expressed linguistically. An example of a rule for estimating productivity can be:

> *If the crew skill level is low and the complexity is high, then the productivity is very low.*

In a Mamdani-type fuzzy rule-based system, the output of each rule is a fuzzy set that matches a linguistic term. For example, Figure 2.6 represents an example of the fuzzy sets of the linguistic terms of the productivity of an activity. The output of the fuzzy inference component in the Mamdani-type fuzzy rule-based system is a fuzzy set that is resulted from the aggregation of the output of different rules. For example, Figure 2.7 represents an example of the output of the fuzzy inference component. The defuzzification component converts this fuzzy set to a crisp value as the final output of the fuzzy rule-based system.

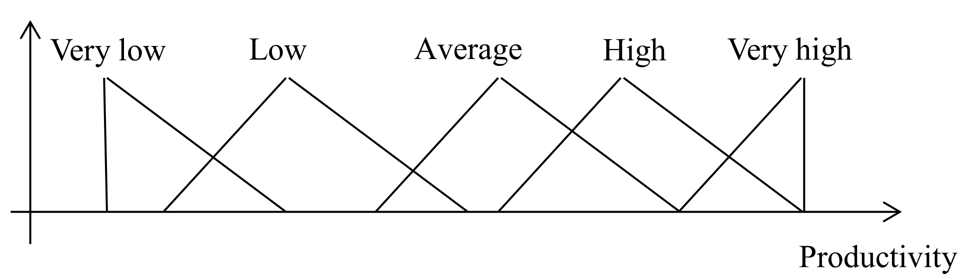


Figure 2.6. An example of the fuzzy sets of the linguistic terms of the productivity of an activity

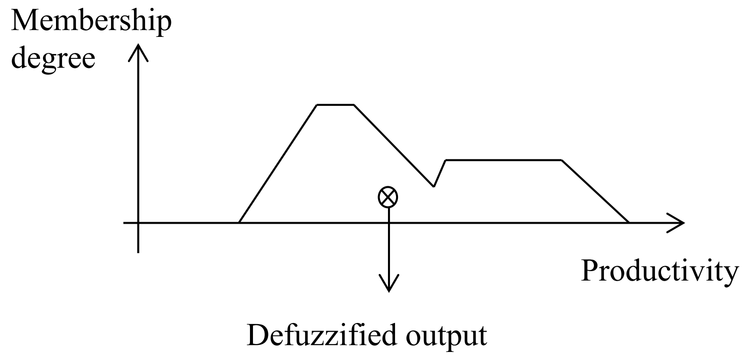


Figure 2.7 An example of the output of the fuzzy inference component in a Mamdani-type fuzzy rule-based system

Although the outputs of fuzzy rule-based systems are presented as crisp values in many situations, some researchers argued that the subjective nature of fuzzy rule-based system suggests that this output will contain subjective uncertainty (Janssen et al. 2010). In the next section, the available approaches for estimating the uncertainty of the output of fuzzy rule-based systems and other prediction models in general will be discussed.

2.4 MODELLING UNCERTAINTY OF PREDICTION

As discussed, prediction models such as ANN, regression, or fuzzy rule-based systems have been used to estimate the activity durations of contraction projects. Therefore, to estimate the uncertainty of activity durations, the uncertainty of the output of such prediction models should be estimated. Any prediction model contains uncertainty (error) in its estimated results. According to Walker et al. (2003), the uncertainty of a prediction model can be categorized as follows:

- *Context uncertainty*: this uncertainty results from uncertain choices in selecting the output and input parameters to be modelled.

- *Model structure uncertainty:* the structure of a prediction model can pose limitations that introduce some uncertainty to a model.
- *Model technical uncertainty:* hardware and software errors can cause uncertain outputs and behaviour from a computer program.
- *Parameter uncertainty:* many prediction models can have parameters that are based on the choice of system modeller. These parameters can affect the uncertainty of prediction.
- *Input uncertainty*: the inputs of a prediction model can have uncertainty in their estimated values.

Model technical uncertainty is not in the scope of this research as it is related to the implementation of a prediction model and can be removed by improving such implementation. However, context uncertainty, model structure, and technical uncertainty are all part of the specifications of a model. Additionally, when a model is trained from experimental data, some uncertainty will be introduced to the model and its outcome due to subjectivity and errors in data.

In a data-rich environment, researchers may divide the data to train and test, using test data, the uncertainty of the model's predictions can be estimated statistically. These approaches are usually based on some assumption regarding the underlying distribution of the probability distribution of uncertainty; however, non-parametric methods have been also proposed for these estimations (Kubisa and Turzeniecka 1996, Manonkian 1986, Pugachev 1984, Sachs 1984).

On the other hand, when available data are limited and it is costly to gain further data, researchers may estimate the uncertainty of model's prediction either analytically or by efficient sample re-use (e.g. cross-validation and the bootstrap). When data are limited, which is usually the case in construction projects, the researcher would like to use the available data for training, with limited data remaining for validation or testing purposes. Analytical approaches are applicable only on specific types of prediction models. Moreover, cross validation and bootstrap outputs have bias and error in their approximated values. Therefore, it is difficult to build accurate probability distributions for representing the uncertainties of the outcomes of many prediction models (Chatfield 2006, Hastie et al. 2001). When enough data for accurate estimation of model uncertainty is not available or the underlying distribution type of the uncertainty is not known, the uncertainty of prediction can be appropriately presented using fuzzy numbers (Baudrit et al. 2006). Some researchers propose methodologies for modelling uncertainty of prediction models using fuzzy set theory (Mauris et al. 2001, Urbanski and Wąsowski 2003, Xia et al. 2000). However, these methods have yet to be applied for representing the uncertainty of prediction models that are predicting durations or productivities of construction activities.

In prediction models that are based on fuzzy rule-based systems no data may be available to allow estimation of the uncertainty of the model. In these situations, it has been suggested that the structure of the fuzzy rule-based system or the defined linguistic terms can be used to represent the uncertainty of the output. Janssen et al. (2010) suggest that the fuzzy output of fuzzy inference component before defuzzification can represent the uncertainty of prediction. Moreover, Roychowdhury and Pedrycz (2001) state that the output of a fuzzy inference component contains uncertainties that will be ignored through defuzzification. However, there are some shortcomings in directly using the output of a fuzzy rule-based system without defuzzification as a representation of output uncertainty; for example, different aggregation methods can be utilized in a fuzzy rule-based system and the shape of the fuzzy output using each method can be quite different. Furthermore, increasing the number of rules of a fuzzy rule-based system can increase the accuracy of the model. However, the undefuzzified output of a fuzzy rule-based system can be even more uncertain as the number of rules increase.

Fayek and Oduba (2005) used the linguistic expression of the output of the fuzzy rule-based system as a method for representing the subjective uncertainty. The output of the fuzzy rule-based system is matched to one of the linguistic terms defined on the output. The linguistic match can be found by either finding the Euclidean distance of the undefuzzified output of the model with the fuzzy sets of linguistic terms; Or by choosing the linguistic term in which the crisp output of the model attains the highest membership degree. This approach of representing the uncertainty of fuzzy rule-based system also has some shortcomings. For example, the uncertainty that is estimated using this approach is not sensitive to the number of rules in the fuzzy rule-based system. Also, using this approach, the output of the model changes from one linguistic term to another and the output does not smoothly change by smooth changes of inputs. Thus, further research is required for representing the uncertainty of the output of fuzzy rule-based systems.

2.5 CONCLUDING REMARKS

A brief review of methods for estimating durations and productivities of construction activities is provided in this chapter. These methods are discussed in two categories. In the first category, activity durations or productivities are estimated using probability distributions or fuzzy numbers. The impact of different factors on productivity or durations of construction activities are implicitly considered in these estimates.

In the second category, a prediction model is developed that can explicitly consider the impact of different factors on durations or productivities of construction activities. These prediction models may be developed from data or by using expert knowledge. In either case, the values estimated using a prediction model contain uncertainties due to the inaccuracy of prediction. These uncertainties are sometimes large in models that are predicting durations or productivities of construction activities. Generally, the uncertainties of prediction models may be presented using probability distributions or fuzzy numbers depending on the availability of data and type of prediction model. However, very limited effort in the area of construction engineering and management has been made for representing these uncertainties. Further research is required to employ or develop methods to represent the uncertainty of models that are predicting productivities or durations of construction activities.

In the next chapter, a methodology for developing data-driven fuzzy rule-based systems for estimating productivities of construction activities is provided. A new approach for estimating the uncertainty of fuzzy rule-based systems using fuzzy numbers is also presented.

2.6 REFERENCES

Abourizk, S. M., & Halpin, D. W. (1992). Statistical properties of construction duration data. *Journal of Construction Engineering and Management, ASCE*, 118(3), 525-543.

Abourizk, S. M., & Sawhney, A. (1993). Subjective and interactive duration estimation. *Canadian Journal of Civil Engineering*, 20(3), 457-470.

Abourizk, S. M., & Wales, R. J. (1997). Combined discrete-event/continuous simulation for project planning. *Journal of Construction Engineering and Management*, *123*(1), 11-20.

Ayyub, B. M., & Haldar, A. (1984). Project scheduling using fuzzy set concepts, *Journal of Construction Engineering and Management*, 110(2), 189-204.

Baudrit, C., Dubois, D., & Guyonnet, D. (2006). Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. IEEE Transactions on Fuzzy Systems, 14(5), 593-608.

Baudrit, C., Dubois, D., & Guyonnet, D. (2006). Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. IEEE Transactions on Fuzzy Systems, 14(5), 593-608.

Beer, M., Ferson, S., & Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. Mechanical Systems and Signal Processing, 37(1), 4-29.

Bertsekas, D. P., & Tsitsiklis, J. N. (2002). Introduction to Probability. Athena Scientific. Belmont, MA.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Chang, C. H., Johnson, N. C., & Cassar, N. (2013). Neural network-based estimates of Southern Ocean net community production from in-situ O2/Ar and satellite observation: a methodological study. Biogeosciences, 11(12), 3279-3297.

Chatfield, C. (2006). Model uncertainty. Encyclopedia of Environmetrics, John Wiley and Sons, Ltd.

Cooper, J. A., Ferson, S., & Ginzburg, L. (1996). Hybrid processing of stochastic and subjective uncertainty data. Risk Analysis, 16(6), 785-791.

Corona-Suárez, G. A., AbouRizk, S. M., & Karapetrovic, S. (2014). Simulation-based fuzzy logic approach to assessing the effect of project quality management on construction performance. Journal of Quality and Reliability Engineering, *1(1), 1-14*.

Corona-Suárez, G. A., Abourizk, S. M., and Karapetrovic, S. (2014), Simulation-Based Fuzzy Logic Approach to Assessing the Effect of Project Quality Management On Construction Performance, *Journal of Quality and Reliability Engineering*, *1(1):1-14*.

Dong, C., Tan, Q., Huang, G. H., & Cai, Y. P. (2014). A dual-inexact fuzzy stochastic model for water resources management and non-point source pollution mitigation under multiple uncertainties. Hydrology and Earth System Sciences Discussions, 11(1), 987-1022.

Fente, J., Knutson, K., & Schexnayder, C. (1999). Defining a Beta Distribution Function for Construction Simulation, Proceedings of *Winter Simulation Conference* , Austin, TX, USA, pp. 1010-1015.

Fente, J., Knutson, K., & Schexnayder, C. (1999). Defining a beta distribution function for construction simulation. Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future, Austin, TX, Volume 2, 1010-1015, ACM.

Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator: L_2 theory. Probability theory and related fields, 57(4), 453-476.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. Journal of the American Statistical Association, 100(470), 680-701.

Gerek, İ. H., Erdis, E., Mistikoglu, G., & Usmen, M. (2014). Modelling masonry crew productivity using two artificial neural network techniques. Journal of Civil Engineering and Management, (ahead-of-print), 1-8.

Gurney, K. (1997). An introduction to neural networks. CRC press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning theory. New York, Springer.

Helton, J. C. (1997). Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. Journal of Statistical Computation and Simulation, 57(1-4), 3-76.

Jahani, E., Muhanna, R. L., Shayanfar, M. A., & Barkhordari, M. A. (2014). Reliability assessment with fuzzy random variables using interval Monte Carlo simulation. Computer‐Aided Civil and Infrastructure Engineering, 29(3), 208-220.

Jahani, E., Muhanna, R. L., Shayanfar, M. A., and Barkhordari, M. A. (2014), Reliability Assessment with Fuzzy Random Variables using Interval Monte Carlo Simulation, *Computer‐Aided Civil and Infrastructure Engineering*, *29*(3), 208-220.

Janssen, J. A. E. B., Krol, M. S., Schielen, R. M. J., Hoekstra, A. Y., & de Kok, J. L. (2010). Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models. Ecological Modelling, 221(9), 1245-1251.

Karshenas, S., & Feng, X. (1992). Application of neural networks in earthmoving equipment production estimating. Computing in Civil Engineering and Geographic Information Systems Symposium, Dallas, TX, 841-847, ASCE.

Kubisa, S. and Turzeniecka, D. (1996), Evaluation of Some Approximated Methods of Measurement Uncertainty Estimation. Proceedings of *Third International Symposium on Methods and Models in Automation and Robotics*, Miedzyzdroje, Poland, 2, pp. 537–42.

Kubisa, S., & Turzeniecka, D. (1996). Evaluation of some approximated methods of measurement uncertainty estimation. Proceedings of the Third International Symposium on Methods and Models in Automation and Robotics, Miedzyzdroje, Poland, Vol. 2, 537-42.

Liu, B. (2009). Theory and practice of uncertain programming, Vol. 239, New York, Springer.

MacCrimmon, K. R., & Ryavec, C. A. (1964). An analytical study of the PERT assumptions. Operations Research, 12(1), 16-37.

Mamdani, E. H. (1977). Application of Fuzzy Logic to Approximate Reasoning using Linguistic Synthesis. *IEEE Transactions On Computers*, 100(12), 1182-1191.

Mamdani, E. H. (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Transactions on Computers, 100(12), 1182-1191.

Manoukian, E. B. (1986). *Modern concepts and theorems of mathematical statistics*. Berlin, Springer.

Mauris, G., Lasserre, V., & Foulloy, L. (2001). A fuzzy approach for the expression of uncertainty in measurement. Measurement, 29(3), 165-177.

Mauris, G., Lasserre, V., and Foulloy, L. (2001), A Fuzzy Approach for the Expression of Uncertainty in Measurement, Measurement, *29*(3), 165-177.

Moselhi, O., Hegazy, T., & Fazio, P. (1991). Neural networks as tools in construction. Journal of Construction Engineering and Management, 117(4), 606-625.

Parzen, E. (1962). On estimation of a probability density function and mode. The Annals of Mathematical Statistics, 1065-1076.

Pedrycz, W. (1994). Why triangular membership functions? Fuzzy sets and Systems, 64(1), 21-30.

Pedrycz, W., & Gomide, F. (2007). Fuzzy systems engineering: toward human-centric computing. John Wiley & Sons, Hoboken, NJ.

Peterson, C., & Miller, A. (1964). Mode, median, and mean as optimal strategies. Journal of Experimental Psychology, 68(4), 363.

Pugachev, V. S. (1984). Probability theory and mathematical statistics for engineers. Elsevier.

Roychowdhury, S., & Pedrycz, W. (2001). A survey of defuzzification strategies. International Journal of intelligent systems, 16(6), 679-695.

Sachs, L. (1984). Applied statistics. Springer.

Scott, D. W. (1979). On optimal and data-based histograms. Biometrika, 66(3), 605-610.

Shaheen, A. A., Fayek, A. R., & AbouRizk, S. M. (2005). A framework for integrating fuzzy expert systems and discrete event simulation. Proceedings of of the Construction Research Congress, San Diego, CA, 1373-1383.

Shimazaki, H., & Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. Neural computation, 19(6), 1503-1527.

Singh, A., Walker, D. D., Minsker, B. S., & Valocchi, A. J. (2010). Incorporating subjective and stochastic uncertainty in an interactive multi-objective groundwater calibration framework. Stochastic Environmental Research and Risk Assessment, 24(6), 881-898.

Smith, S. D. (1999). Earthmoving productivity estimation using linear regression techniques. Journal of Construction Engineering and Management, 125(3), 133-141.

Song, L., & AbouRizk, S. M. (2008). Measuring and modeling labor productivity using historical data. Journal of Construction Engineering and Management, 134(10), 786-794.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics, (1), 116-132.

Urbanski, M. K., & Wąsowski, J. (2003). Fuzzy approach to the theory of measurement inexactness. Measurement, 34(1), 67-74.

Weiler, H. (1965). The use of incomplete beta functions for prior distributions in binomial sampling. Technometrics, 7(3), 335-347.

Xia, X., Wang, Z., & Gao, Y. (2000). Estimation of non-statistical uncertainty using fuzzy-set theory. Measurement Science and Technology, 11(4), 430.

Zadeh, L. A. (1965). Fuzzy Sets, *Information And Control*. 8(3), 338-353.

Zadeh, L. A. (2008). Is there a need for fuzzy logic? *Information Sciences,*178(13), 2751-2779.

Zeng, Z., Xu, J., Wu, S., & Shen, M. (2014). Antithetic method‐based particle swarm optimization for a queuing network problem with fuzzy data in concrete transportation systems. Computer‐Aided Civil and Infrastructure Engineering, 29(10), 771-800.

Zhang, H., Tam, C. M., & Li, H. (2005). Modeling uncertain activity duration by fuzzy number and discrete-event simulation. *European Journal of Operational Research*, 164(3), 715-729.

CHAPTER 3 A FRAMEWORK FOR DEVELOPING DATA-DRIVEN FUZZY RULE-BASED SYSTEMS FOR PREDICTING CONSTRUCTION LABOUR PRODUCTIVITY

3.1 INTRODUCTION

The productivity can be measured as the ratio of the quantity of the output to the input. On the other hand, some researchers measure the productivity as the ratio of the input to the output (Park et al. 2005). The first approach for defining the productivity (i.e. quantity of the output to the input) is used in this dissertation. In construction projects, productivity can be measured at different levels such as company level, project level, activity level, or a certain discipline (e.g. pipers or electricians). Labour productivity in construction projects is usually referred to the productivity at the activity level, where the input is measured in man-hours and the output is measured as completed quantities (Dozzi and AbouRizk 1993). The unit for the output quantity is defined depending on the type of activity. For example, the output for the structural steel erection or for the welding activity can be measured in tons of erected structural steel or welded diameter inches of the spool, respectively.

Accurate estimation of construction productivity plays an important role for scheduling, estimating and making decisions in construction projects (Sonmez and Rowings 1998). However, various factors impact the productivity of construction projects that make an accurate productivity prediction difficult. Explicit consideration of some of the significant factors on productivity can increase the accuracy of productivity prediction (AbouRizk and Sawhney 1993).

As discussed in Chapter 2 of this dissertation, various models have been developed to predict construction labour productivity. If data are available for both the set of influencing factors and the productivity, machine-learning methods can be used to train a productivity prediction model. Linear regression models and artificial neural networks (ANN) are the most common approaches for developing construction labour productivity prediction models from data. However, the following shortcomings have been identified in models that predict construction labour productivity:

- Linear regression models assume linear relationships among the variables. As a result, they usually have less predictive accuracy compared to ANN models (Tu 1996).
- ANN is powerful in solving complicated problems such as productivity prediction. However, the main criticism to ANN models is that they are black boxes; In other words, one can train an ANN model, which provides good predictions of the productivity, without providing an explanation of the behaviour of the system (Benítez et al. 1997).
- The inevitable error in productivity prediction models results in an uncertainty in the predicted value. This uncertainty is usually considerable regardless of the modelling approach. However, very little attention has been paid to the representation of the uncertainty in construction labour productivity prediction models.

Considering the above shortcomings, the objectives of this chapter is set as followings:

1) To propose a methodology for developing a data-driven construction labour productivity prediction model that is interpretable

2) To provide a methodology for representing the output uncertainty of the predicted productivity

In this research, a fuzzy rule-based system is proposed to provide an interpretable prediction model for predicting construction labour productivity for the following reasons:

- Fuzzy rule-based systems can be expressed with interpretable linguistic rules rather than providing an ANN black box that takes inputs and provides outputs.
- Fuzzy rule-based systems combined with optimization algorithms such as artificial neural networks (ANN) or genetic algorithms (GA) can be trained on historical data.
- Fuzzy rule-based systems are universal approximates, meaning that they are theoretically able to model any function (Wang 1992; Castro and Delgado 1996).

The proposed framework is presented in three main steps to predict construction labour productivity: 1) Identifying factors and collecting data; 2) Developing fuzzy rule-based system; 3) Estimating output uncertainty. These steps are presented in sections 3.2, 3.3, and 3.4, respectively. The practicality of the proposed framework is further illustrated using a real example for estimating the productivity of structural steel erection in Section 3.5. Finally, Section 3.6 presents concluding remarks of this chapter.

3.2 IDENTIFYING FACTORS AND COLLECTING DATA

For developing a data-driven productivity prediction model, enough data should be collected for the influencing factors as well as the productivity. First, an initial list of factors should be developed from the literature. Many researchers analyze and provide the list of influencing factors on construction labour productivity in different areas and contexts of construction projects (Park et al. 2005). The literature often presents factors impacting the construction labour productivity in terms of categories. A list of factors is often defined under each category. Various categories of factors are considered to be important for estimating the productivity of construction activities in the literature. For example, Borcherding and Alarcon (1991) identified seven categories and used 38 sub-factors for each category. Their identified categories are: 1) schedule acceleration, 2) poor coordination, 3) changes, 4) management characteristics, 5) project characteristics, 6) labour and morale, and 7) project location and external conditions. AbouRizk et al. (2001) defined nine categories with 33 sub-factors as following: 1) general project characteristics, 2) site characteristics, 3) labour, 4) equipment, 5) overall project difficulty, 6) general activity, 7) quantity, 8) design, and 9) activity difficulty.

When preparing the initial list of influencing factors, the measurement method for the suggested factors may be also extracted from the literature, if applicable. This is because for many factors in construction projects, a standard measurement method (unit) does not exist. For example, the productivity of material handling for spools can be measured in tons or number of spools. After preparing the initial list of factors, the list may be updated based on the context of the activity or project for which the productivity prediction model is being developed. The possible updates of the list include removing irrelevant factors, updating or adding possible measurement methods, or adding new factors. Structured surveys have been proposed in the literature to develop or to update the list of factors based on expert judgment (e.g. Dai and Goodrum 2012).

Many of the factors affecting the duration of construction projects are qualitative such as complexity and workers' skill levels. These factors cannot be assessed numerically and should be assessed using a scale or as linguistic terms. The researcher may opt to break a qualitative factor into more detailed influencing factors to decrease subjectivity and to potentially increase the accuracy of data collection. For example, spool complexity is a qualitative factor that can be expressed in terms of more objective factors such as number of spools, weight and length of the spools and number of welds. However, including more objective factors may not totally remove the qualitative element of the complexity that may not be possible to be fully presented through the objective factors. A rating scale is defined to collect data for qualitative factors. The rating scales should be defined as tangible as possible to reduce the subjectivity. For this purpose, predetermined rating scales that define different aspects of one factor using more detailed sub-factors are proposed by Awad (2012). The predetermined rating scale tries to improve consistency of the model when having different experts for different inputs of the model. This is due to the fact that employing the predetermined rating scales provides interviewees (i.e. experts) a more consistent understanding of the defined scales and thus decreasing the subjectivity in the inputs.

After finalizing the list of factors and measurement methods, data should be collected for the identified influencing factors as well as the output productivity. Data collection is usually the most time consuming and costly step for developing construction labour productivity prediction models. Various sources of information may be used to collect data for training a prediction model. For example, the following tools may be utilized for this purpose:

- Timesheet systems: Computerized timesheet systems are designed for tracking the labour-hours of construction activities.
- Time study and on-site observations: Performing time studies and on-site observations is a common tool for estimating the duration of construction activities.
- Computerized drawings and plans for quantity surveying: Computer-based drawings are great resources for extracting quantities and specifications for construction activities. For example, CAD isometric drawings may be used to find the specifications and quantities of pipe spool (Rokni and Fayek 2010) and steel fabrication (Song and AbouRizk 2008).
- Questionnaires and interviews: In circumstances in which data of the factors have not been recorded and is not feasible to collect data using on-site observations, questionnaires and surveys can be employed to collect required data.
- Available databases and documents: In recent years, companies have developed databases and documents containing useful information that can be employed to extract various factors and productivities for past projects.

- Past schedules: Companies' schedules of past projects may also be used to mine historical data regarding the duration of activities.

By completing the data collection stage, a prediction model can be developed for predicting construction productivity based on the identified influencing factors. The development of this prediction model is discussed in the following section.

3.3 DEVELOPING PRODUCTIVITY PREDICTION MODEL

A fuzzy rule-based system is trained from the collected data in this stage for developing an interpretable productivity prediction model. In recent years, different methodologies for developing and tuning fuzzy rule-based systems from data are proposed. Two of the most promising approaches are neuro-fuzzy systems (Jang 1993), and genetic fuzzy systems (Chiu 1994). In these methods, artificial neural networks (ANN) or genetic algorithm (GA) are employed to learn the rules or tune the parameters of fuzzy rule-based system. However, in high dimensional problems, it is often very complicated to learn the if-then fuzzy rules (i.e. to find the optimal antecedent and consequent of the rules and number of rules in the fuzzy rule-based systems) using ANN or GA. Thus, it is recommended that a rule-based system is defined using another method and ANN or GA only be used for optimizing (i.e. tuning) the parameters of the fuzzy rule-based system (Abraham 2001, Cordon et al. 2001). Fuzzy C-means (FCM) has been specifically proposed for constructing antecedents and consequents of fuzzy if-then rules when facing high dimensional datasets (Tsekoura 2005, Nuovo et al. 2007, Ahmad and Pedrycz 2011).

Considering various factors impacting the productivity of construction activities, the dataset containing the data of influencing factors of construction activities is often high dimensional. Therefore, FCM is employed as an effective methodology for developing fuzzy rule-based systems from this high dimensional dataset (Tsekouras 2005, Nuovo et al. 2007, Ahmad and Pedrycz 2011).

This section is organized as follows: In Section 3.3.1, FCM approach is explained. In Section 3.3.2, an integrated approach for selecting input features and developing a fuzzy rule-based system using FCM is presented. Section 3.3.3 discusses the methodology for interpreting the developed model. Finally, in Section 3.3.4 the membership functions of the fuzzy rule-based are optimized for optimum accuracy.

3.3.1 FCM for Developing Fuzzy Rule-Based Systems

FCM is a clustering method based on fuzzy set theory. In crisp (non-fuzzy) clustering, each instance (i.e. data point) is either a member or not a member of a cluster. In fuzzy clustering, each instance has a degree of membership in each cluster. The FCM algorithm is first developed by Dunn (1973) and enhanced later by Bezdek (1981). This algorithm finds c cluster centres $\{c_1, c_2, \ldots, c_c\}$ where points further to the cluster center have a lesser membership degree compared with points closer to the cluster center. The membership degree of each point x_i in cluster j is presented as w_{ij} in the partition matrix. The standard function for calculating the membership function is presented in Equation 3.1.

$$w_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d(x_i,c_j)}{d(x_i,c_k)}\right)^{\frac{2}{m-1}}} \tag{3.1}$$

In Equation 3.1, *d(a,b)* represents the distance from *a* to *b*. Moreover, *m* is a real number greater than 1 and indicates the degree of fuzziness of the solution. The cluster centers are calculated according to the center of points weighted by their membership degree in that cluster as illustrated in Equation 3.2, where *n* is the number of data points.

$$c_k = \frac{\sum_{i=1}^{n} w_{ik}^{m} * x_i}{\sum_{i=1}^{n} w_{ik}^{m}} \tag{3.2}$$

In summary, the FCM algorithm is based on the following procedure:

1) Randomly initialize the partition matrix, *W*
2) Calculate the cluster centers using Equation 3.2.
3) Update the partition matrix, *W', based* on Equation 3.1.
4) If the difference between the *W* and *W'* is less than a threshold, stop; Otherwise, assign *W'* to *W* (*W*←*W'*) and go to step 2.

Each of the clusters developed using FCM algorithm represents a rule in the fuzzy rule-based system. The fuzzy membership functions for the antecedents and consequents of the rules are formed based on the projection of the developed clusters on the input and output space (Delgado et al. 1997, Nauck and Kruse 1999). Thus, both the number of rules and the number of membership functions on the inputs and output are equal to the number of clusters in the developed fuzzy rule-based system. The number of clusters may be optimized to maximize the accuracy of the fuzzy rule-based system. However, the interpretability is lower as the number of clusters increase. For achieving a higher interpretability, the number of clusters should be considered as a small number such as 3, 5, or 7.

The membership functions generated based on projection of fuzzy clusters on different inputs and output of the model contain some ripples. The intensity of these ripples depends on the fuzzification coefficient (variable m in equations 3.1 and 3.2). For higher values of m, the rippling effect is higher. The commonly used value for m in FCM equals to 2, but can be subjected to optimization (Pedrycz and Gomide 2007). Figure 3.1 illustrates an example of three membership functions generated using FCM for m=2.

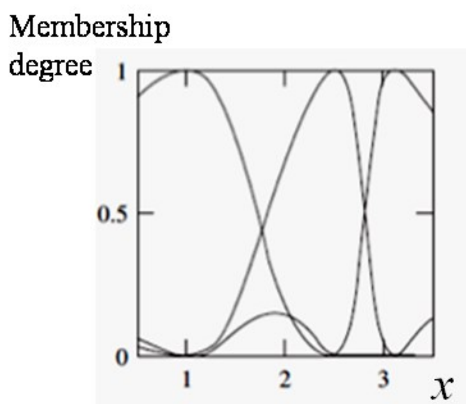


Figure 3.1 An example of three membership functions generated using FCM and the rippling effect for m=2 (Adapted from Pedrycz and Gomide 2007)

As illustrated in Figure 3.3, the shape of the projected membership functions are not interpretable; A membership function that can be related to a linguistic term such as *low* or *high* is expected to be unimodal. To interpret the fuzzy rule-based system, the membership functions can be approximated with a standard membership function such as Gaussian, trapezoidal or triangular membership functions to achieve higher interpretability. Gaussian membership functions are employed in the proposed approach because of their smooth shapes. A Gaussian membership function, *G,* can be defined according to Equation 3.3.

$$\mu\ (x)=e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{3.3}$$

The linguistic terms can be assigned to the membership functions generated using the FCM method. For example, when 5 clusters are used in the FCM algorithm, the membership functions developed for the productivity can be labeled as very low, low, average, high, and very high (Figure 3.2).

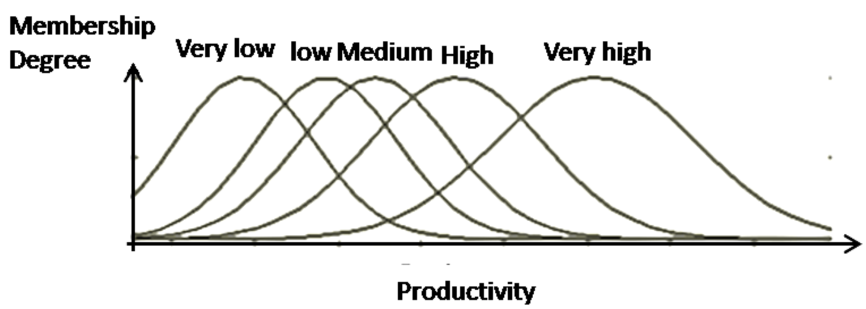


Figure 3.2 Gaussian membership functions developed for productivity

(left to right: very low, low, average, high, and very high)

Furthermore, for higher interpretability, the first and the last membership functions (modelling concepts such as very low or very high temperature) can be adjusted according to sigmoidal membership function with two parameters represented in Equation 3.6, where parameter *c* in this equation is the crossover point with membership of 0.5, and parameter *a* controls the slope of the sigmoidal membership function.

$$\mu(x) = \frac{1}{1+e^{-a(x-c)}} \tag{3.4}$$

For example, Figure 3.3 depicts two sigmoidal functions having parameter *a* as 1 and -1. If parameter *a* has a negative value, the function is open to the left and is appropriate for modelling concepts such as very low. Otherwise, the function is open to the right and is good for modelling concepts such as very high. These two parameters (a and c) are subjected to optimization, as will be discussed later.

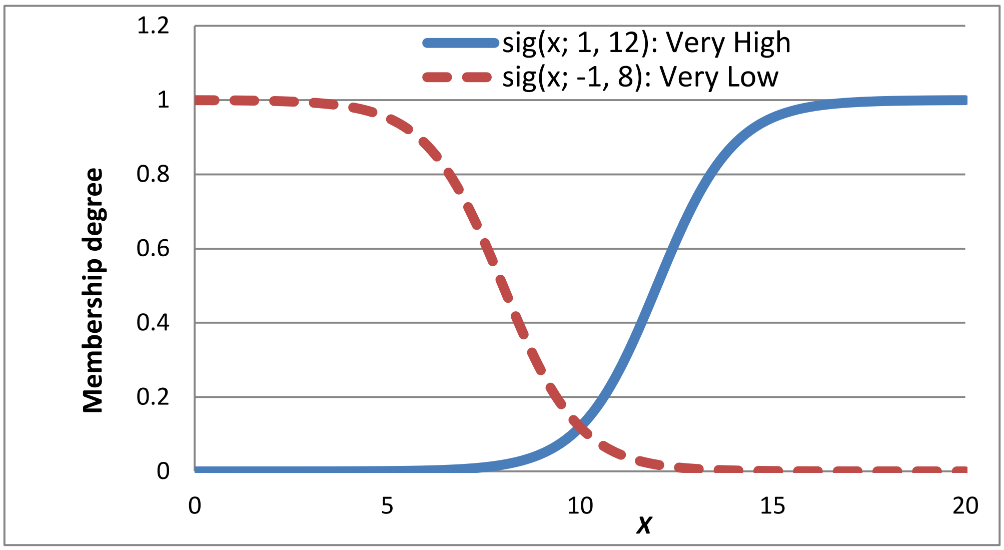


Figure 3.3 Example of sigmoidal membership functions

In some situations, the membership functions can be too close to define two different linguistic terms. Different similarity measures are proposed in the literature to provide a degree of closeness of two fuzzy sets (e.g. Chen and Linkens 2001). In this research project, expert judgment is used to decide if two fuzzy sets are too close for defining two distinct linguistic terms. This is because the expert should be able to interpret the linguistic terms to each of the membership functions at the end. When two membership functions are identified to be very close, the two membership functions can be merged to form a new membership function; the new fuzzy membership function can be calculated by taking the average of the two original membership functions (Chen and Linkens 2001) (e.g. Figure 3.4 (b)). This average value is calculated based on fuzzy arithmetic.

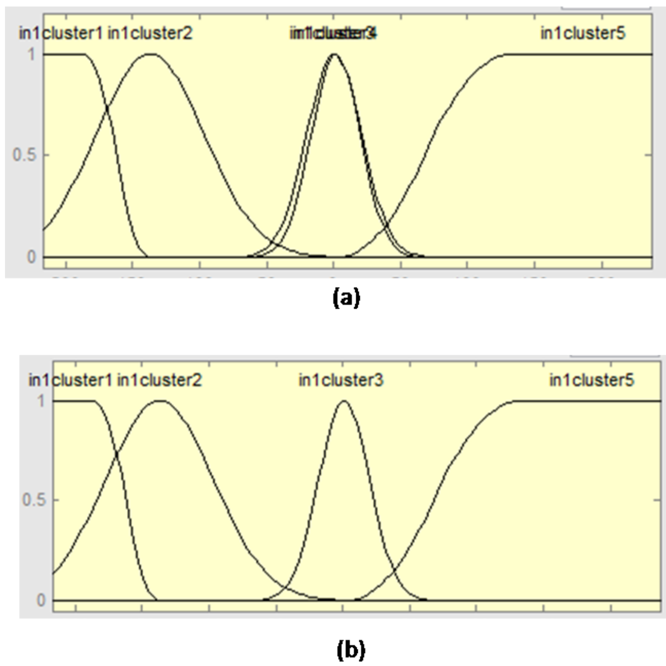


Figure 3.4 Membership function defined on an input variable (a) two membership functions that are very close (b) merging very close membership functions

Using this proposed approach, the rules in the fuzzy rule-based system can be expressed linguistically. However, the proposed approach is only used for interpreting the input and output of the fuzzy rule-based system. On the other hand, for calculating the output of the model, the original membership functions of input parameters are maintained to preserve model accuracy. This is because using the Gaussian or sigmoidal membership functions in the developed fuzzy rule-based system will result in the sparse fuzzy rule-based system. When the fuzzy rule-based system is sparse, for some of the input observations, no rule may be fired. On the other hand, classical fuzzy inference systems are designed to deal with complete or dense fuzzy rule-based systems in which the rule premises completely cover the input space. Keeping the original fuzzy membership function with the ripples will resolve the problem of sparseness, as each membership function will cover a wider spectrum of the input space. For the output membership functions however, the approximated Gaussian membership functions are used, as they do not impact to the problem of sparseness. However, because the parameters of the Guassian membership functions are approximated for the output membership functions, these parameters are subject to optimization as discussed in the following section.

Using the FCM approach, an initial fuzzy rule-based system can be developed. However, the developed fuzzy rule-based system should be improved in two aspects: 1) optimizing input features; 2) optimizing the parameters of the fuzzy rule-based system. These aspects are discussed in the following subsections.

3.3.2 Optimizing Input Features

In many high dimensional datasets, using a subset of features can enhance the realization of a fuzzy model. Thus, feature selection methods have been proposed for developing more accurate fuzzy rule-based systems from high dimensional datasets (Ahmad and Pedrycz 2011).

Feature selection is used to reduce the dimensionality and to increase the performance of the initial fuzzy rule-based system. In machine learning with a fixed number of training samples, the predictive power decreases as the number of features increases (curse of dimensionality). As a result, reducing the number of features and selecting a certain number of features for the input can increase the predictive power of the model (Hughes

1968). Feature selection reduces the dimensionality of data by selecting only a subset of measured features to create a model. There are three main approaches for feature selection:

- Filter methods: Filter methods select features without optimizing the performance of a predictor. These methods can be combined with a search method to rank different subsets of features based on a heuristic metric. Filter methods are the cheapest approach for feature selection.
- Wrapper methods: Wrapper methods use the performance of a learning machine trained using a given feature subset. The wrapper methods are based on the search algorithms. They search through different subsets of features and evaluate the machine learning performance on each feature subset.
- Embedded methods: Embedded methods perform feature selection in the process of training. These methods are specific to given learning machines.

In developing a productivity prediction model, one may try a number of feature selection methods to find the most suitable one. Both commercial and free programs exist specifically for selecting features using filter methods. One of the simplest yet powerful methods among filter methods is Pearson correlation coefficient. Although, this approach simply rank the features based on correlation, it has been proven to perform very well in experiment (Guyon et al. 2006, NIPS 2003 workshop). Pearson's correlation coefficient between two variables, X and Y is defined as the covariance of the two variables, *cov(x,y),* divided by the product of their standard deviations , $\sigma_x \sigma_y$ (Equation 3.3).

$$r_{xy} = \frac{cov(X,Y)}{\sigma_x \sigma_y} \tag{3.5}$$

Ranking the subset of features can be also performed using the correlation based feature selection (CFS) (Hall, 1998). CFS ranks feature subsets based on heuristic evaluation function as indicated in Equation 3.6.

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}} \tag{3.6}$$

In this equation, M_s is the heuristic metric for the feature subset *S* with *k* features, $\overline{r_{cf}}$ is the mean correlation of the features and output, and $\overline{r_{ff}}$ is the mean correlation of the features in the class with each other.The hypothesis used in CFS is that a good feature subset should contain features that are highly correlated with the output, yet uncorrelated with each other. The numerator in Equation 3.4 indicates how predictive the subset of features is. The nominator indicates how much redundancy exists among the features of the subset.

The wrapper methods for feature selection are based on the search algorithms, searching through different subsets of features and evaluating the machine learning performance on each feature subset. Ahmad and Pedrycz (2011) proposed wrapper methods specifically for fuzzy rule-based systems developed with FCM. For a feature subset, a new fuzzy rule-based system is developed using FCM. The performance of fuzzy rule-based system is calculated based on the root mean square error (RMSE) of the fuzzy rule-based system. The RMSE can be calculated according to Equation 3.5. In this Equation, *Y* is a vector of *n* predictions and *Y'* is a vector of actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i')^2} \qquad (3.5)$$

In the proposed wrapper feature selection method, the heuristic optimization algorithm finds the feature subset to minimize the RMSE of the fuzzy rule-based system developed with FCM. Genetic Algorithm (GA), inspired from the process of natural selection, is a powerful heuristic optimization algorithm that can be applied for feature selection. GA has been proposed as an effective strategy for feature selection by various researchers (e.g. Yang and Honavar 1998, Leardi 2000, Li et al. 2011, Guo et al. 2011). The underlying mechanism of basic GA is as follows:

1. A random population of chromosomes is generated where each chromosome represents a possible solution to the problem. The most basic method of representing a chromosome is a bit string (arrays of 0s and 1s).
2. A fitness value is calculated for each of the chromosomes according to a fitness function. The fitness function is defined for each problem representing the degree of optimality, adaptation, or quality of a solution.
3. A subset of the chromosomes is selected from the initial population of chromosomes. In the selection process, the chromosomes with better fitness values have higher probabilities to be selected.
4. Crossover and mutation operators are applied on the selected population to develop a new generation of population. Crossover operator is employed to develop child chromosomes by combining sections of parent chromosomes. The section can be developed based on one-point, two points or more points in the parent chromosomes. For example, Figure 3.1 represents a two-point crossover

for generating the children of two parents. Mutation operator is the random substitution of one or more values in a chromosome with another value.

5. Repeat the process until the termination condition is reached. Typical termination conditions are for example reaching a satisfactory fitness level, maximum number of generation, or not being able to produce better results in the successive iterations.
6. The final solution is the chromosome with the best fitness value.

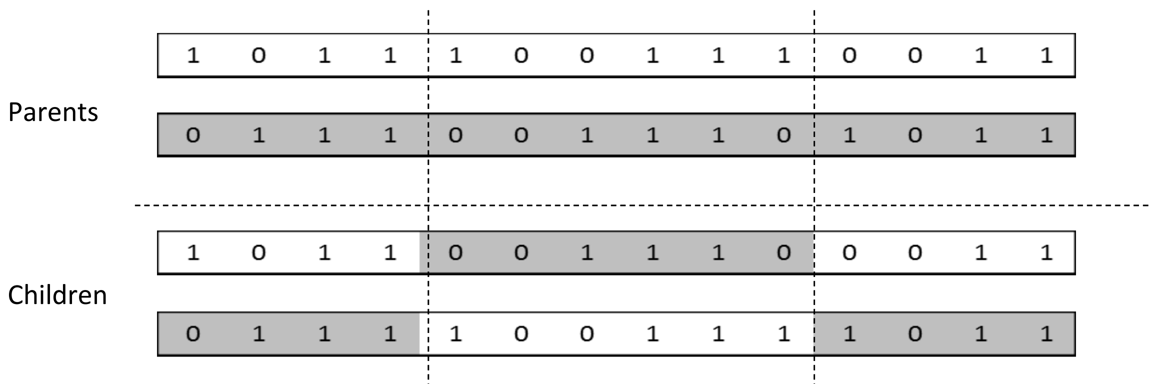


Figure 3.5 Crossover of chromosomes in GA process

When applying GA for feature selection, each chromosome represents a subset of features. The chromosome is represented as a bit string; each bit is corresponded with a feature. A bit equal to 1 indicates that the feature is included in the feature subset while a bit equal to 0 indicates the feature is not included. The fitness function for feature selection in the fuzzy rule-based system is proposed to be equal to RMSE of the model on data (Ahmad and Pedrycz 2011). Thus, a feature subset with a smaller RMSE has a better fitness. Using the explained approach, the feature subset of the model for developing the fuzzy rule-based system with FCM can be optimized.

3.3.3 Tuning the Parameters of Fuzzy Rule-Based System

The parameters of the fuzzy rule-based system can be optimized to increase the accuracy of prediction. Two of the most successful methods for tuning the parameters of fuzzy rule-based systems are ANN and GA. The fusion of ANN and fuzzy rule-based systems is referred as neuro-fuzzy systems. In neuro-fuzzy systems, the main idea is to benefit from the powerful learning capability of ANN, yet provide an interpretable fuzzy if-then rule rather than a black box. Various structures of neuro-fuzzy systems are provided in the literature (Abraham 2001). GA has been also extensively applied to learn the rules or optimize the parameters of fuzzy rule-based systems. The fusion of GA and fuzzy rule-based systems is referred as genetic fuzzy systems (Cordon 2011).

A fuzzy rule-based system contains several parameters that are subject to optimization. The parameters of the membership functions, the rules, or defuzzification strategy are some examples of these parameters. However, two conflicting objectives should be considered for decision regarding the parameters to be optimized (Cordon 2001):

1) Completeness: a complete search space that includes all of the parameters of the fuzzy rule-based system is more likely to provide the optimal solution.
2) Optimization efficiency: the optimization process is performed faster and more efficient in a smaller search space.

Therefore, there is a trade-off between efficiency and completeness for selecting the parameters of fuzzy rule-based systems to be optimized. As a result, different designs are proposed for fuzzy rule-based optimization problems. A comprehensive literature review of different genetic fuzzy systems is presented by Cordon (2001). A review of different neuro-fuzzy systems is also provided by Sahin et al. (2012). By employing one of the available approaches, the performance of the fuzzy rule-based system can be enhanced.

In this proposed approach, GA is used to optimize the parameters of the output membership functions in the fuzzy rule-based system. This is because; the Gaussian membership functions in the developed fuzzy rule-based system are approximated from the projected FCM clusters as discussed previously. Thus, by tuning the parameters of these membership functions, better model accuracy may be achieved. Each of the parameters output membership functions are presented as a gene in the chromosomes of GA. Thus, using 5 clusters in FCM and two parameters for each Gaussian membership function, each chromosome includes 10 genes of type double. This is because, for each of the features and the output, we have 5 membership functions, each with 2 parameters that need to be optimized. Similar to the feature selection methodology that is discussed in Section 3.3.2, RMSE is used as the fitness function in the GA algorithm.

This is the last step in the proposed methodology for developing an interpretable productivity prediction model from data. However, any productivity prediction model developed with any approach contains inaccuracy (i.e. error) in the predicted results. Thus, the productivity estimated by a productivity prediction model should be represented as an uncertain variable. In the next section, a novel methodology is proposed for estimating the output uncertainty of data-driven fuzzy rule-based systems.

3.4 ESTIMATING OUTPUT UNCERTAINTY

A fuzzy rule-based system includes rules that are expressed using linguistic terms. The fuzzy inference component in the fuzzy rule-based system employs the membership functions of the linguistic terms to perform reasoning and to calculate the output. The output of a fuzzy inference component is a fuzzy set that is defuzzified by the defuzzification component. However, representing the output of the fuzzy rule-based system as a crisp (defuzzified value) ignores the inaccuracy of the model resulted from fuzzy reasoning as discussed in Chapter 2.

Some researchers proposed that the output of a fuzzy rule-based system should be presented as a fuzzy number due to the use of fuzzy membership functions in its reasoning methodology (Janssen et al. 2010, Roychowdhury and Pedrycz 2001). For estimating the output uncertainty of data-driven fuzzy rule-based systems, the performance of the model on actual data can be employed. In this section, the approach proposed by Pedrycz and Gomide (2007) is employed to represent the output of the fuzzy rule-based- system using fuzzy numbers.

In the proposed approach, firstly, the output of fuzzy rule-based system is calculated as a crisp value. Then, two parameters a and b are estimated to define a fuzzy number A around the *output* to represent the uncertainty of the prediction model, tri(*output*-*a*,*output*,*output*+*b*).

The value of the parameter a and b are optimized based on the overall performance of the fuzzy rule-based system on available data. The "theory of justifiable granularity" is employed to optimizing the parameters of fuzzy numbers on experimental data (Pedrycz and Gomide 2007). Theory of justifiable granularity provides two criteria for estimating the parameters of fuzzy numbers:

- The fuzzy numbers have to reflect the experimental data to the highest extent. Thus, the sum of the membership degree of the experimental data in the output fuzzy number should be maximized.
- Second, the fuzzy numbers should have a well defined semantic and should be as specific as possible. Thus, the objective function is to maximize the specificity of the fuzzy number.

In the case of a fuzzy rule-based system, assume for each input set ($input_i$), the output fuzzy number A_i is estimated as tri(yp_i-a,yp_i,yp_i+b), where yp_i is the crisp output of the fuzzy rule-based system using COA (center of area) defuzzification method. The actual value of the output is referred as y_i.. According to the first criteria, the sum of the membership degree of each data point, y_i, in its predicted output fuzzy number A_i should be maximized. Figure 3.6 represents an example of the membership degree of y_i in A_i, $\mu_{Ai}(y_i)$. This objective can be expressed as $max(\sum_i \mu_{Ai}(y_i))$, where μ_{Ai} is the membership function of fuzzy number A_i . At the same time, according to the second criteria, the specificity of the output fuzzy numbers, A_i should be also maximized. According to Pedrycz and Gomide (2007), $e^{-(a+b)}$ can be maximized to maximiz the specificity of the triangular membership function, tri(yp_i-a, yp_i, yp_i+b).

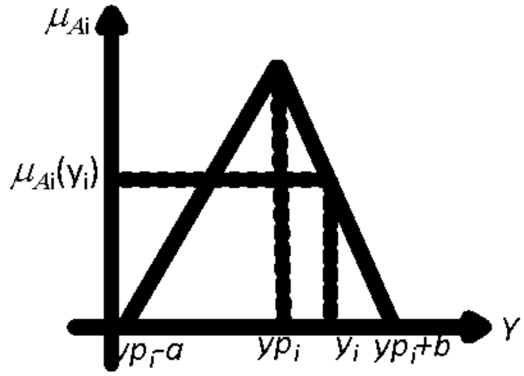


Figure 3.6 An example of the membership degree of actual value y_i in the output of fuzzy rule-based system A_i for the input set, input$_i$

In order to optimize the two conflicting objectives of maximizing membership grades and specificity, multi objective optimization procedures can be employed. These procedures are able to develop a trade-off curve for the bi-criteria optimization problem. However, since we are interested in defining a fuzzy number by considering a balance between these two objective functions, developing trade-off curve that usually depends on expert to make the final decision is not necessary. Thus, a single objective function can be defined by combining the two objective functions (Stewart 1992). Additive approaches of combining these two objective functions are not appropriate in this problem because when one objective will go toward infinity, the other will go toward 0. However, additive objective functions will assume this answer acceptable while it does not provide a balance between specificity and coverage. As proposed by Pedrycz and Gomide (2007), considering the multiplication of specificity and coverage is a legitimate objective function to this problem. Therefore, the objective can be defined according to Equation 3.8.

$$Max((\sum_i \mu_{Ai}(x_i)) * (\sum e^{-(a+b)}) \quad (3.8)$$

The parameter a in Equation 3.7 for the estimated output membership functions is optimized based on the objective function in Equation 3.8 using GA. Therefore, GA employs chromosomes with a single gene, where the gene has a type double and is greater than zero. The fitness function of GA is defined according to Equation 3.8.

Using the proposed approach, a fuzzy number representing the uncertainty of the output of a prediction model is estimated based on the performance of the model on data. The practicality of the proposed methodology to develop a fuzzy rule-based system for productivity prediction model and to estimate the output uncertainty as a fuzzy number is illustrated in the following section.

3.5 A CASE STUDY OF MODULE ASSEMBLY YARD

Modules are preassembled units that make construction of oil-sands refining facilities fast and easy. Pipe spool modules are usually assembled offsite, in the module assembly yard, which is usually located near a pipe spool fabrication shop, and are then transported to the site. The modules are assembled using prefabricated components such as structural steel frames, cables, and pipe spool components that are fabricated in the pipe spool fabrication shop (Taghaddos et al. 2009).

The process of assembling a module in a module assembly yard includes different activities. The productivities of these activities are impacted by numerous factors. The objective of this section is to illustrate the practicality of the proposed framework of developing a productivity prediction model based on fuzzy rule-based system and representing the output as a fuzzy number. For illustrative purposes, the fuzzy rule-based system is developed for the structural steel erection, which is one of the first activities in the module assembly process. The productivity prediction model is developed based on the data from a module fabrication company near Edmonton, AB. In Section 3.5.1, the process for identifying factors and data collection is explained. Section 3.5.2 discusses the development of fuzzy rule-based system and estimating the output uncertainty.

3.5.1 Identifying Factors and Collecting Data

In order to define the appropriate list of factors that are influencing the productivities of activities of the module assembly process, the categories and factors proposed by AbouRizk et al. (2001) are used initially. This is due to the fact that AbouRizk et al. (2001) mainly focused on pipe installation and welding productivities, which are closely related to the activities in the module assembly yard. However, this list is enhanced based on some other published research papers as well as the company's documents. A three-layer approach (using category, factor, and sub-factor) is employed for a more organized representation of factors. For example, average crew size and peak crew size are defined as two sub-factors. These two sub-factors are presented under a more general factor referred as crew characteristics. The crew characteristics are also considered under the category of activity difficulty as proposed by AbouRizk et al. (2001).

Various references in the literature regarding construction productivity are reviewed to enhance the list of factors including Hanna et al. (2008), Hanna et al. (2002), Hanna et al.(1999), Korde et al. (2005), Moselhi and Khan (2012), Soekiman et al. (2011), Thomas and Sakarcan (1994), Chan and Kumaraswamy (1997), and Hsieh et al. (2004).

For example, “season” is defined as a factor by AbouRizk et al. (2001). Moselhi and Khan (2012) defined four factors that can potentially better quantify weather: (1) temperature; (2) humidity; (3) precipitation; and (4) wind speed. Therefore, weather is defined as a general factor and temperature, humidity, precipitation, wind speed, and season are defined as sub-factors of weather.

Using the above approach, an initial list of factors is prepared. This list contains eight general categories: 1) general project; 2) site; 3) tradesperson and foreman; 4) equipment; 5) project difficulty; 6) general activity; 7) activity difficulty; and 8) activity design and quantities. Activity design and quantity may contain different numbers of sub-factors, ranging from 4 to 75 factors depending on the activity. This is because the complexity and factors affecting different activities in the module assembly yard are extremely different. For the rest of the categories, a total number of 26 general factors with 77 sub-factors are defined. For each sub-factor, a measuring unit is proposed. Figure 3.7 indicates the list of factors developed under the first category, general project.

Category	general Factor	Sub-Factor	Unit of Measure	Measure Type	Reference
1. General project	1. Superintendent	Superintendent's historical performance	Productivity	Real number	AbouRizk, 2001
		Capability of superintendent to	Very low/Low/Moderate/High/Very high		Kadir et al. 2005
	2. Project manager	Project manager's historical performance	Productivity	Real number	AbouRizk, 2001
	3. Client	Timely response to questions and inquiries	Strongly agree/Agree/Neither agree nor disagree/Disagree/Strongly disagree	Qualitative rating	Chan and Kumaraswamy
		Average time to response to questions and inquiries	#Days	Real number	Dai and Goodrum, 2012
		Quality of coordination with client	Very low/Low/Moderate/High/Very high		
		Years with client	#Years	Real number	Hanna et al. 1999
	4. Consultants	Average time for verifying progress claims	#Days		Kadir et al. 2005
		Quality of coordination with consultants	Very low/low/moderate/high/very high	Qualitative rating	Kadir et al. 2005
	5. Subcontractor	Timely finishing of tasks by sub-contractors	Strongly agree/Agree/Neither agree nor disagree/Disagree/Strongly disagree	Qualitative rating	Kadir et al. 2005
		Average amount of delays by sub-contractors	#Days	Real number	Kadir et al. 2005
		Quality of coordination with subcontractors	Very low/Low/Moderate/High/Very high	Qualitative rating	Kadir et al. 2005
	6. Contract	Type of contract	Guaranteed maximum price/Lump sum/Unit price/Cost plus/Cost-reimbursable alternative/Integrated	Categorical	Kadir et al. 2005
	7. Economy	Economical conditions	Very poor/Poor/Moderate/Good/Very good	Qualitative rating	Kadir et al. 2005
		Inflation rates	%Rate	Real number	Kadir et al. 2005
		Bank interest rates	%Rate	Real number	Kadir et al. 2005
	8. Safety	Recordable incident rate(RIR)	%Rate	Real number	Hoonaker et al. 2005
		Lost workday case incident rate(LIR)	%Rate	Real number	Hoonaker et al. 2005
		Having adequate safety plans and practices for the project	Strongly agree/Agree/Neither agree nor disagree/Disagree/Strongly disagree	Categorical	AbouRizk, 2001

Figure 3.7 General factors and sub-factors identified under the category of general project

A structured survey is used to update and to improve the list of factors using 15 respondents to the designed surveys. The interviewees are general managers, managers, and a random selection of foremen, and labourers of the module assembly yard. The random selection is based on performing the interviews at random days and random selection of interviewees during breaks. The complete survey is presented in Appendix A of this dissertation. The interview surveys are used to modify the list of factors, and the researcher relied on the actual data to identify the final impact of factors on productivity. The interviewees were asked questions regarding the impact, availability of data, alternative measuring method, and new sub-factors in these interview surveys. A field is designed in the survey document for each of these questions. These fields are explained in the following:

- Impact indicates the degree that a factor is believed to be important in impacting the productivity. A 5- or 7-point Likert scale is usually used in the literature for measuring the impact (Albaum 1997). A 5-point scale is employed in this research as it provides easier interpretation by the interviewees, where 1 is very low or no impact, 2 is low, 3 is moderate, 4 is high, and 5 is very high.
- Availability of data indicates how easily and accurately the data can be obtained for a specified sub-factor for the past 5 years of module assembly projects in the company. Before the pilot survey, only two options for data availability were considered, which was either available or not available. After the pilot survey, the researcher realized that the interviewees (module assembly yard managers and superintendents) prefer to express the availability of data in a more subjective manner. For example, they would like to say some data are available, or we can look at the documents to derive some data for that factor. Therefore, the surveys are developed as a 5-point Likert scale to rate availability of data, where 1 is very low or no data, 2 is low, 3 is moderate, 4 is high, and 5 is very high. The actual availability of data is realized during the process of data collection. However, the rate of data availability is used as an indication to further investigate about the factor. If the impact rate of a factor was higher than 2, but the data availability is rated less than 4, the interviewee were asked regarding the possible ways to better and more accurately assess this factor. This

was performed by asking them to suggest some alternative measurement methods or other sub-factors.

- Alternative measurement methods are presented as a field in the survey that allows defining a different unit of measurement for the sub-factors than the chosen one.
- New sub-factors are defined as another field in the survey allowing the interviewees to suggest new sub-factors for each category. New factors suggested by interviewees in the surveys are included in the list of factors.

The above approach is used to develop a list of factors before starting data collection. After finalizing the list of factors, qualitative and quantitative factors are identified. For the qualitative factors, the approach suggested by Awad (2012) is employed to define predetermined rating scales for each of the qualitative factors. In this approach, a list of clarifying factors is defined that is: (1) concise to be easily readable by the interviewee; (2) decrease the subjectivity and improves the clarity of a factor; and (3) cover different aspects of the more general factor that is not covered by other sub-factors.

Previous research in this area (e.g. Poveda and Fayek 2009, Tsehayae and Fayek 2014) and expert inputs are employed to develop the factors for the predetermined scales for the factors. In order to define the statements that match different combinations of scale ratings to a rating of the factors in a Likert scale, the rules are developed using expert's knowledge based on the relative importance of factors of the predetermined rating scales (Awad 2012). One may suggest collecting data on each of the sub-factors separately and later combining them instead of defining a predetermined scale as suggested by Awad

(2012). However, the number of factors to be surveyed would increase significantly which may decrease the quality of inputs by experts. Furthermore, in the case that the defined sub-factors are also subjective; it is usually difficult for the experts to provide specific values for each sub-factor. The approach for defining values for the sub-factors is tested in a pilot experiment with one of the factors, complexity of spool. Two different aspects of complexity of spools that are suggested by the experts of the company are considered: (1) congestion of pipes and equipment in the module, and (2) complexity of configuration of pipes. In the pilot experiment, the expert rated both of these factors always equal for all of the module cases (more than 150 modules). This pilot test was conducted with the experts that were among the most knowledgeable and collaborative managers in the company. Therefore, this similarity of results is not due to the ignorance but rather due to the fact that the two aspects of module complexity, although are different, are closely related and it is very difficult for the interviewee to rethink about these aspects separately and rate them accordingly. On the other hand, the expert prefers to rate the complexity of the module as one entity. However, other factors related to the complexity that can be quantified numerically, such as the number of levels in a module and module dimensions, are considered as separate factors in this study.

After defining the list of factors and their measurement method, the data are collected for the identified list of factors for the past projects of the company. For this purpose, a database is developed in Microsoft Access. The data collected using different sources are stored as different tables in this database. The following sources are used to collect data:

- The reports from the tracking system of the company regarding the dates of completing activities, progress quantity and hours, number of RFI and etc.
- The payrolls of the company containing the actual hours spending on each activity, the position of the workers working on each activity, the dates, and etc.
- The surveys which are developed for collecting data on qualitative factors as well as factors for which data was not available in the database; For example, surveys are used for complexity of module and skill levels associated with each worker's positions.
- The weather data extracted from Environment Canada website (Environment Canada).

Structured Query Language (SQL) (Ramakrishnan 2003) is used to extract the dataset of the module activity process in the required format from the developed Access database. Due to the deficiencies of the available data in the tracking system of the company, data are not collected for some of the factors that are considered in the initial list. At the end, data was collected for 45 influencing factors for 13 activities in the module assembly process. Moreover, some of the factors are considered as context variables as the data are limited to 5 years' worth of project of one company. Thus, the values for those variables do not have enough variations to be used in training the fuzzy rule-based system. By excluding those factors, the final number of factors that can be used for training is reduced to 33. After finalizing the data collection, the fuzzy rule-based system for predicting the productivity of module assembly process is developed as presented in the following section.

3.5.2 Developing the Fuzzy Rule-Based System

The collected data for the structural steel erection of the module assembly yard is used to develop the productivity prediction model for the structural steel erection activity. The collected dataset for the structural steel erection activity includes 971 data points. 90% of this data points are used for training and 10% are used for testing. The approach proposed in Section 3.3 is used to develop the fuzzy rule-based system.

For developing the initial model, MATLAB function called "FCM" is employed. FCM is able to generate clusters from data. As discussed, each cluster is considered as a rule. Thus, estimated cluster centers from FCM are used to calculate the degree that each rule is fired in the fuzzy rule-based system according to Equation 3.1. The number of clusters in FCM is considered equal to 5 to achieve a high interpretability. In the developed fuzzy rule-based system, maximum method is used for aggregation, and COA is employed for defuzzification process.

Using the above approach, a fuzzy rule-based system is developed in MATLAB for predicting the productivity of structural steel erection. The RMSE (Equation 3.5) of the prediction model is equal to 0.42 tons/man-hour when all of the input factors are considered in the model. To improve this performance, two feature selection methods are employed and compared:

1) A filter technique is employed using a data mining Software in Java, called Weka 3 (Weka 3: Data Mining Software in Java), based on correlation coefficient (Hall, 1998). Six out of the 33 features are selected using this

approach. The model developed with the features selected by Weka has an RMSE equal to 0.31 tons/man-hour.

2) A wrapper technique is employed using the genetic algorithm (GA). GA is applied to select the features as discussed in Section 3.3.3. The method is implemented in MATLAB using the available function for Genetic Algorithm (GA). Five features out of 33 features are selected by this approach and the RMSE of the developed model with these four features is equal to 0.23.

Thus, the features selected using the GA is used to develop the fuzzy rule-based system to achieve higher model accuracy. These features are: average of skill level of workers, average number of revisions of design documents, average of temperature, the maximum of wind speed (Gust), and the complexity of module. The developed fuzzy rule-based system is interpretable containing only 5 linguistically defined rules. In order to provide an interpretable rule-based system, MATLAB function “Genfis3” is employed. “Gnefis3” generates a fuzzy rule-based system and approximate the projected membership functions of each cluster as a Gaussian membership function. The fuzzy toolbox in MATLAB allows viewing and modifying the fuzzy rule-based system (Figure 3.8).

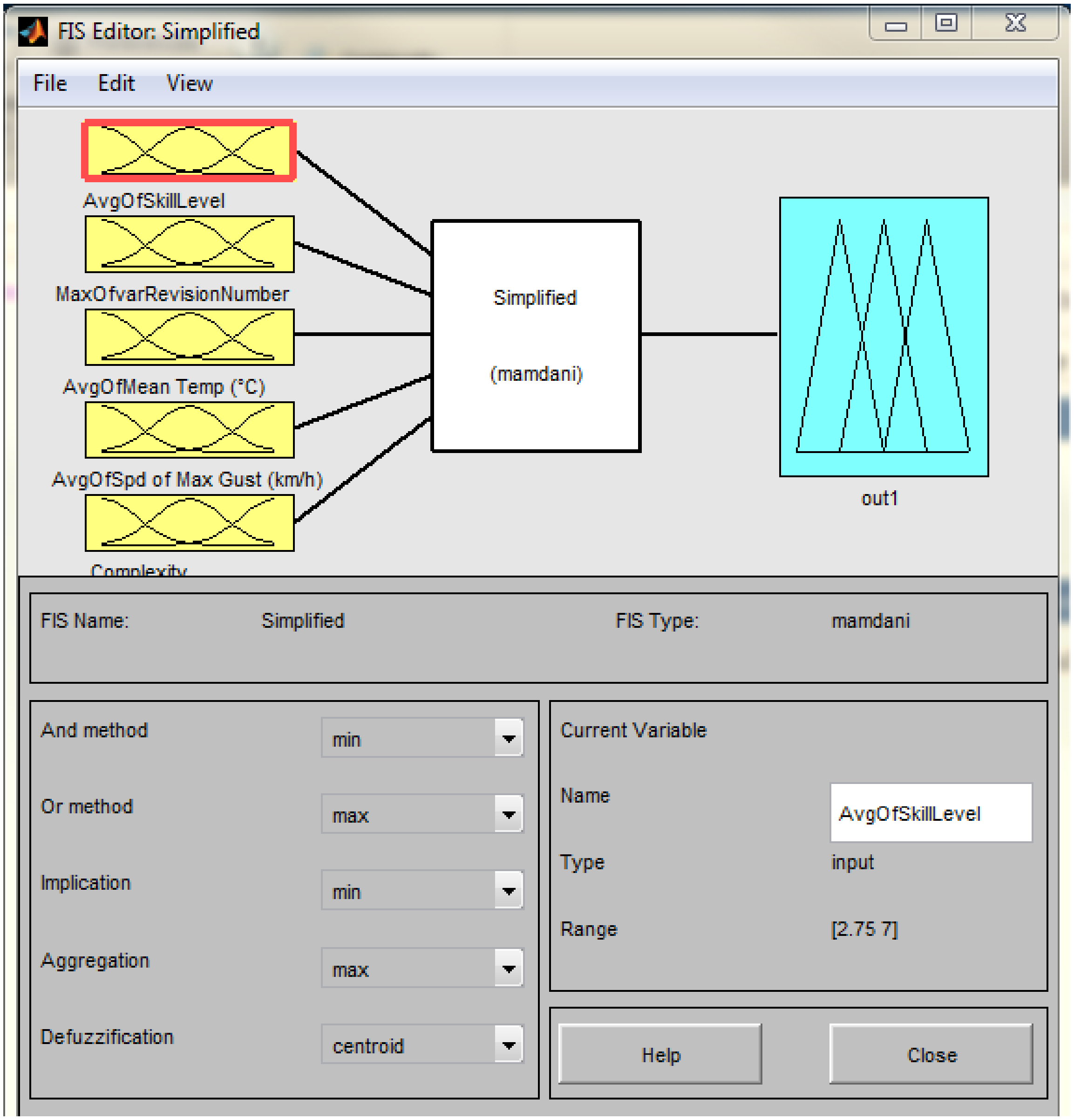


Figure 3.8 View of a the developed fuzzy rule-based system in MATLAB fuzzy toolbox

Expert judgment is used to assign the linguistic terms to each of the inputs as discussed in this Chapter. After modifying the rules, we can view the rules as linguistic descriptions of the model. For example one of the rules of this fuzzy rule-based system is:

If the average skill level of workers is *poor*, and the average number of revisions of design documents is *high*, and the average of temperature is *medium*, and the maximum of wind speed is *low*, then the productivity is low.

For further improvement of the performance of the fuzzy rule-based system, the parameters of the output membership functions are further tuned using GA. Both of the parameters of the Gaussian membership function (Equation 3.6) are tuned to increase the model accuracy. After tuning the parameters of the fuzzy rule-based system, the RMSE of the model becomes 0.18 tons/man-hour. Finally, the values of *a* and *b* for representing the output of the fuzzy rule-based system as a fuzzy number are optimized with GA according to the procedure discussed in section 3.4. The optimized values for *a* and *b* are 0.20 and 0.12 tons/man-hour respectively. Thus, the output fuzzy number is skewed toward smaller values.

The performance of the proposed approach for estimating the productivity of structural steel erection is compared with the performance of an ANN model. The ANN productivity prediction model is developed in MATLAB using the same data and same input features that are employed in the developed fuzzy rule-based system. The developed ANN model contains 3 layers and 10 nodes in the middle layer. The “fitnet” function and “train” function to develop this productivity prediction model are employed from MATLAB ANN toolbox. The default parameters of the ANN toolbox are used in training ANN productivity prediction model. The default training method in this toolbox is Levenberg-Marquardt which is a commonly used method for training neural network models. The RMSE of the developed ANN model is 0.15 tons/man-hour, which is slightly smaller than the RMSE of the developed fuzzy rule-based system. However, the advantage of the proposed framework over ANN method is in its interpretability which comes with the price of a slight loss of accuracy in this example. Generally, there is usually a trade-off between interpretability and accuracy of prediction models.

3.6 CONCLUDING REMARKS

In this chapter, a comprehensive framework for developing interpretable construction productivity prediction models based on fuzzy rule-based systems is proposed. The proposed approach uses fuzzy C-means clustering to develop an initial rule-based system. Furthermore, genetic algorithm is used to optimize the features and parameters of the fuzzy rule-based system. A novel approach has been also developed to model the output of the fuzzy rule-based system as a fuzzy number to represent the uncertainty of prediction.

The practicality and effectiveness of the proposed approach has been illustrated through a real example of estimating the productivity of structural steel erection in a module assembly yard. This case study shows that explicitly considering the factors impacting the productivity decreases the uncertainty of the predicted productivity compared with representing the productivity as one single distribution that implicitly represents the impact of different factors. However, the uncertainty of the prediction should be appropriately represented to provide a realistic estimation. In the future, other methods for developing interpretable productivity prediction models (such as neuro-fuzzy systems) can be employed and compared with the proposed methodology.

Representing the output uncertainty of the fuzzy rule-based system as proposed in this chapter is especially important when performing further analysis with the predicted productivity. For example, the estimated fuzzy productivity may be used to estimate the activity duration as a fuzzy number. This activity duration can become the input to a project network or a simulation model, which will be then used for scheduling and making decisions. Fuzzy activity durations can be analyzed in event-based construction simulation models using fuzzy discrete event simulation frameworks. In the next section, the fuzzy discrete-event simulation framework for processing fuzzy activity duration is presented.

3.7 REFERENCES

Abourizk, S. M., & Sawhney, A. (1993). Subjective and interactive duration estimation. *Canadian Journal of Civil Engineering*, 20(3), 457-470.

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.

Abraham, A. (2001). Neuro fuzzy systems: State-of-the-art modeling techniques. *Connectionist models of neurons, learning processes, and artificial intelligence* Proceedings, pp. 269-276, Springer Berlin Heidelberg.

Ahmad, S. & Pedrycz, W. (2011). Feature and Instance selection via cooperative PSO. Systems, Man, and Cybernetics (SMC) Proceedings, pp. 2127-2132, IEEE.

Ahmad, S. S. S., & Pedrycz, W. (2011). Feature and Instance selection via cooperative PSO. Systems, Man, and Cybernetics (SMC) Proceedings, pp. 2127-2132, IEEE.

Albaum, G. (1997). The Likert scale revisited. *Journal-Market research society*, *39*, 331-348.

Awad, A. (2012). Intelligent contractor default prediction model for surety bonding in the construction industry. Doctoral dissertation, University of Alberta.

Babuska, R. (1998). Fuzzy modeling for control. Kluwer Academic Publishers, Boston, MA.

Baraldi, P., Razavi-Far, R., & Zio, E. (2011). Bagged ensemble of Fuzzy C-Means classifiers for nuclear transient identification. *Annals of Nuclear Energy*, *38*(5), 1161-1171.

Benítez, J. M., Castro, J. L., & Requena, I. (1997). Are artificial neural networks black boxes? *Neural Networks, IEEE Transactions on*, *8*(5), 1156-1164.

Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers Norwell, MA, USA.

Borcherding, J. D. , and Alarcón, L. F. (1991). Quantitative effects on construction productivity. The Construction Lawyer, 11 (1), 36–48.

Castro, J. L., & Delgado, M. (1996). Fuzzy systems with defuzzification are universal approximators. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 26*(1), 149-152.

Chan, D. W., & Kumaraswamy, M. (1997). A comparative study of causes of time overruns in Hong Kong construction projects. *International Journal of Project Management*, *15*(1), 55-63.

Chen, M. Y., & Linkens, D. A. (2001). Rule-base self-generation and simplification for data-driven fuzzy models. *The 10th IEEE International Conference on Fuzzy Systems Proceedings,* Vol. 1, pp. 424-427, IEEE.

Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy systems*, *2*(3), 267-278.

Cordón, O. (2011). A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. *International Journal of Approximate Reasoning*, *52*(6), 894-913.

Cordón, O., Herrera, F., Gomide, F., Hoffmann, F., & Magdalena, L. (2001). Ten years of genetic fuzzy systems: current framework and new trends. *IFSA World Congress and 20th NAFIPS International Conference Proceedings,* Vol. 3, pp. 1241-1246, IEEE.

Dai, J., & Goodrum, P. M. (2012). Generational differences on craft workers' perceptions of the factors affecting labour productivity. Canadian Journal of Civil Engineering, 39(9), 1018-1026.

Delgado, M., Gómez-Skarmeta, A. F., & Martin, F. (1997). A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling. IEEE Transactions on *Fuzzy Systems,* *5*(2), 223-233.

Di Nuovo, S., Palesi, M., & Catania, V. (2007). Multi-objective evolutionary fuzzy clustering for high-dimensional problems. Fuzzy Systems Conference Proceedings, 2007. FUZZ IEEE International (pp. 1-6). IEEE.

Dozzi, S. P. & AbouRizk, S. M. (1993). Productivity in construction. Institute for Research in Construction, National Research Council, Ottawa, ON, Canada

Dubois, D., & Prade, H. (1992). The semantics of fuzzy "if… then…" rules. Fuzzy Approach to Reasoning and Decision-Making, pp. 3-16, Springer Netherlands.

Government of Canada. (2014). Environment Canada. Retrieved from http://www.ec.gc.ca/

Grzegorzewski, P., & Mrówka, E. (2005). Trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, *153*(1), 115-135.

Grzegorzewski, P., & Mrówka, E. (2007). Trapezoidal approximations of fuzzy numbers—revisited. *Fuzzy Sets and Systems*, *158*(7), 757-768.

Guo, J., White, J., Wang, G., Li, J., & Wang, Y. (2011). A genetic algorithm for optimized feature selection with resource constraints in software product lines. Journal of Systems and Software, 84(12), 2208-2221.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. (2006). Feature extraction, foundations and applications. Springer.

Hall, M. A. (1999). Correlation-based feature selection for machine learning. Doctoral dissertation, The University of Waikato.

Hanna, A. S., Chang, C. K., Sullivan, K. T., & Lackney, J. A. (2008). Impact of shift work on labor productivity for labor intensive contractor. *Journal of Construction Engineering and Management*, *134*(3), 197-204.

Hanna, A. S., Peterson, P., & Lee, M. J. (2002). Benchmarking productivity indicators for electrical/mechanical projects. *Journal of construction engineering and management*, *128*(4), 331-337.

Hanna, A. S., Russell, J. S., Nordheim, E. V., & Bruggink, M. J. (1999). Impact of change orders on labor efficiency for electrical construction. *Journal of Construction Engineering and Management*, *125*(4), 224-232.

Hsieh, T. Y., Lu, S. T., & Wu, C. H. (2004). Statistical analysis of causes for change orders in metropolitan public works. *International Journal of Project Management*, *22*(8), 679-686.

Hughes, G.F. (1968). On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14 (1): 55–63.

J. C. Dunn (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics,* 3, 32-57.

Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, *23*(3), 665-685.

Johanyák, Z. C., Tikk, D., Kovács, S., & Wong, K. W. (2006). Fuzzy rule interpolation Matlab toolbox-FRI toolbox. *IEEE International Conference on Fuzzy Systems, 2006*, 351-357, IEEE.

Koczy, L. T., & Hirota, K. (1997). Size reduction by interpolation in fuzzy rule bases. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *27*(1), 14-25.

Korde, T., Li, M., & Russell, A. D. (2005). State-of-the-art review of construction performance models and factors. Proceedings of the Construction Research Congress: Broadening Perspectives, San Diego, CA.

Leardi, R. (2000). Application of genetic algorithm–PLS for feature selection in spectral data sets. *Journal of Chemometrics*, *14*(5-6), 643-655.

Li, S., Wu, H., Wan, D., & Zhu, J. (2011). An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems*, *24*(1), 40-48.

Machine Learning Group at the University of Waikato (2014), Weka 3: Data Mining Software in Java, Retrieved from http://www.cs.waikato.ac.nz/ml/weka/

Moselhi, O., & Khan, Z. (2012). Significance ranking of parameters impacting construction labour productivity. *Construction Innovation: Information, Process, Management*, *12*(3), 272-296.

Nauck, D., & Kruse, R. (1999). Obtaining interpretable fuzzy classification rules from medical data. *Artificial intelligence in medicine*, *16*(2), 149-169.

NIPS 2003 workshop (2003), Seventeenth Annual Conference on Neural Information Processing (NIP) Systems, Retrieved from http://clopinet.com/isabelle/Projects/NIPS2003/.

Park, H. S., Thomas, S. R., & Tucker, R. L. (2005). Benchmarking of construction productivity. *Journal of Construction Engineering and Management*, *131*(7), 772-778.

Poveda, C. A., & Fayek, A. R. (2009). Predicting and evaluating construction trades foremen performance: fuzzy logic approach. *Journal of Construction Engineering and Management*, *135*(9), 920-929.

Ramakrishnan, R., Gehrke, J., & Gehrke, J. (2003). *Database management systems.* Vol. 3, New York: McGraw-Hill.

Reddy, C. S., & Raju, K. V. S. V. N. (2009). An improved fuzzy approach for COCOMO's effort estimation using gaussian membership function. Journal of software, 4(5), 452-459.

Roychowdhury, S., & Pedrycz, W. (2001). A Survey of Defuzzification Strategies. *International Journal of Intelligent Systems*, *16*(6), 679-695.

Sahin, S., Tolun, M. R., & Hassanpour, R. (2012). Hybrid expert systems: A survey of current approaches and applications. Expert Systems with Applications, 39(4), 4609-4617.

Setnes, M., & Roubos, H. (2000). GA-fuzzy modeling and classification: complexity and performance. *IEEE Transactions on Fuzzy Systems*, *8*(5), 509-522.

Soekiman, A., Pribadi, K. S., Soemardi, B. W., & Wirahadikusumah, R. D. (2011). Factors relating to labor productivity affecting the project schedule performance in Indonesia. *Procedia Engineering*, *14*, 865-873.

Sonmez, R., & Rowings, J. E. (1998). Construction labor productivity modeling with neural networks. Journal of Construction Engineering and Management, 124(6), 498-504.

Stewart, T. J. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, *20*(5), 569-586.

Taghaddos, H., AbouRizk, S. M., Mohamed, Y., & Hermann, U. (2009). Integrated simulation-based scheduling for module assembly yard. *Proceedings of the 2009 ASCE Construction Research Congress (CRC),* 1270-1279.

Thomas, H. R., & Sakarcan, A. S. (1994). Forecasting labor productivity using factor model. *Journal of Construction Engineering and Management*, *120*(1), 228-239.

Tikk, D., & Baranyi, P. (2000). Comprehensive analysis of a new fuzzy rule interpolation method. *IEEE Transactions on Fuzzy Systems*, *8*(3), 281-296.

Tikk, D., Baranyi, P., Gedeon, T. D., & Muresan, L. (2001). Generalization of the rule interpolation method resulting always in acceptable conclusion. *Tatra Mountains Math. Publ*, *21*, 73-91.

Tsehayae, A. A., & Fayek, A. R. (2014). Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects. *Canadian Journal of Civil Engineering*, *41*(10), 878-891.

Tsekouras, G. E. (2005). On the use of the weighted fuzzy c-means in fuzzy modeling. Advances in Engineering Software, 36(5), 287-300.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, *49*(11), 1225-1231.

Wang, L. X. (1992). Fuzzy systems are universal approximators. *Proceedings of International Conference on Fuzzy Systems*, IEEE, 1163-1170.

Wong, K. W., Tikk, D., Gedeon, T. D., & Kóczy, L. T. (2005). Fuzzy rule interpolation for multidimensional input spaces with applications: A case study. *IEEE Transactions on Fuzzy Systems*, *13*(6), 809-819.

Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. *Feature extraction, construction and selection*, 117-136, Springer US.

Yeh, C. T. (2008). On improving trapezoidal and triangular approximations of fuzzy numbers. *International Journal of Approximate Reasoning*, *48*(1), 297-313.

CHAPTER 4 FUZZY DISCRETE EVENT SIMULATION FOR CONSTRUCTION APPLICATIONS[1]

4.1 INTRODUCTION

Discrete event simulation (DES) is used for construction scheduling, sensitivity analysis to consider different operational strategies, and to calculate construction project estimates. Traditional DES can handle probabilistic input values (stochastic uncertainty). On the other hand, fuzzy discrete event simulation (FDES) is a discrete event simulation approach that can handle fuzzy numbers in its time dependant input variables (e.g., activity durations). As discussed in Chapters 2 and 3 of this dissertation, fuzzy numbers can represent the uncertainty of the activity durations of construction projects when facing lack of data, or subjectivity due to the linguistic expression of knowledge.

In traditional DES, fuzzy numbers cannot be used as an input to the simulation, they must either be defuzzified or converted to a probability distribution. Defuzzification converts the fuzzy number to a crisp number and thereby disregards the uncertainty in the simulation inputs. Conversion of fuzzy numbers to probability distributions is controversial since the nature of these uncertainties is different. Probability distributions represent randomness of a variable, while fuzzy numbers represent the uncertainty in a variable resulting from a lack of precise knowledge or linguistic expression of that variable. Furthermore, there is no commonly accepted method of converting fuzzy

[1] Parts of this chapter is submitted to the journal of Automation in Construction and is published in the Proceedings of International Conference of Fuzzy Systems Association Congress, pp.1102–1106.

numbers to probability distributions (Guyonnet et al. 2003). Therefore, FDES framework for dealing with fuzzy uncertainty in construction discrete event simulation models is required.

Although FDES can greatly benefit the simulation of construction management applications, the methodology for processing fuzzy numbers in FDES is challenging. The fundamental differences between fuzzy numbers and probability distributions introduce new challenges to the engine of the FDES. One of the main challenges is the possible impact of different fuzzy ranking methods on FDES outputs (Anglani et al. 2000, Perrone et al. 2001, Zhang et al. 2005). The time of events and the simulation time are represented with fuzzy numbers in FDES. Therefore, fuzzy ranking is used to determine the order of the events and the advancement of the simulation time in FDES. However, the fuzzy numbers to be ranked in the simulation may be overlapping and challenging to rank or order. Different ranking methods may produce different results of ranking and hence different simulation outcomes (Perrone et al. 2001). Different approaches to fuzzy ranking and updating the simulation time are suggested for FDES in the literature; however, there is no consensus or comprehensive study on the impact of these various approaches on simulation models. Furthermore, current approaches to fuzzy ranking and updating the simulation time in FDES have the problem of either underestimation or overestimation of the simulation time, which will be illustrated in this chapter.

The objective of this chapter is, therefore to: 1) illustrate the shortcomings of available fuzzy ranking approaches in updating the simulation time of FDES; 2) provide a new methodology for FDES that can accurately update the simulation time; and 3) illustrate the practicality of FDES for construction management using practical examples.

The rest of this Chapter is structured as follows: Section 4.2 elaborates on the differences between FDES and DES. In Section 4.3, the impacts of various ranking methods on FDES are discussed. Section 4.4 proposes a new method for updating the simulation time in FDES to eliminate the problem of overestimation or underestimation of the simulation time. Section 4.6 explains the implementation of the proposed FDES methodology. Section 4.5 compares the performance of the proposed FDES methodology with previous approaches using a project network example. Section 4.6 provides a real case study of a tunneling operation to illustrate the practical application of the proposed FDES methodology. Finally, conclusions are provided in Section 4.7.

4.2 DES VERSUS FDES

In systems modelling, if the state of a system continuously changes over time, it is called a continuous system. If the state of a system changes at discrete points of time, it is called a discrete system, which is the case for discrete event simulation. An event is defined as the occurrence of a change of state in a discrete system. For example, an event may be defined for the completion of an activity in DES. DES is based on the scheduling of events. At any point in time, *TNOW*, new events may be generated for future occurrence based on the simulation logic. The time of occurrence of an event, the event time, is calculated by adding the delay time of an event, *D*, to the current simulation time, *TNOW*.

$$Event\ time = TNOW + D \tag{4.1}$$

For example, if all logical conditions for starting an activity are satisfied at time *TNOW* in DES, the event for completion of the activity is generated for future occurrence with the delay time equal to the duration of the activity. The events generated for future occurrences are managed in an event list in DES. For example, Figure 4.1(a) represents an event list with three events, where *TNOW*=4. The simulation time advances in DES by finding the event with the smallest event time. The next event in this event list is Event 2, as it has the smallest event time. The simulation time will advance to the time of the chosen event, and the event will be removed from the event list. Therefore, in Figure 4.1, *TNOW* will advance to time 6. Simulation terminates when all events in the event list have been executed and no additional events are listed.

TNOW=4

Event ID	Event Time
2	6
3	7
6	8

(a)

Advancing the simulation time ⇨

TNOW=6

Event ID	Event Time
3	7
6	8

(b)

Figure 4.1 Event list in DES (a) event list at time TNOW=4 (b) event list at time TNOW=6

The uncertain inputs (e.g., durations of activities) of DES are represented with probability distributions. Samples of the input probability distributions are used to calculate the event times and to perform the simulation. Therefore, the output of a single simulation run is a sample value of the output. To obtain a meaningful output, DES is replicated numerous times to estimate the statistical properties (e.g. mean and variance) of the output.

FDES is a discrete event simulation in which time variables in the simulation (e.g., inter-arrival times, times between failures, activity durations) are fuzzy numbers instead of probability distributions. FDES is able to model uncertainty resulting from vagueness, subjectivity, lack of knowledge, or imprecision. Stemming from the major differences between fuzzy and probability theory, there are major differences between the FDES and the traditional DES. Whereas sampling can be performed on probability distributions in a DES, it cannot be performed on fuzzy numbers in a FDES.

Similar to DES, the time of an event in FDES is calculated based on the delay time of an event plus the simulation time, *TNOW*. However, the delay time (which is the input to the simulation) is a fuzzy number in FDES; therefore, fuzzy arithmetic must be used to calculate the event time. Fuzzy arithmetic can be performed on fuzzy numbers using the alpha-cut method (Dubois 1980). The alpha-cut of the output of a function g at the level of α can be calculated using Equation 4.2. In this equation, the input arguments are alpha-cuts of $A_1 \ldots A_n$, which are intervals. Therefore, arithmetic of intervals is performed to calculate the output interval. The output of function g can be reconstructed from its alpha-cuts at all alpha levels. For this purpose, the alpha-cuts are aggregated based on the representation theorem, which states that each fuzzy set can be reconstructed from its alpha-cuts according to Equation 4.3. In this equation, $\mu_{B_\alpha}(x)$ represents the membership function of the interval B_α. Therefore, $\mu_{B\alpha}(x)$ is 1 if $x \in B_\alpha$ and is 0 if $x \in B_\alpha$.

$$B_{\alpha}=g(A_{1\alpha},\ldots,A_{n\alpha})B_{\alpha} = g(A_{1\alpha},\ldots,A_{n\alpha}) \tag{4.2}$$

$$\mu_B(x)=\sup_{\alpha\in[0,1]}\alpha\,\mu_{B_\alpha}(x)\mu_B(x) = \sup_{\alpha\in[0,1]}\alpha\mu_{B_\alpha}(x) \tag{4.3}$$

Equation 4.4 indicates the calculation of *Event time* in FDES, where $\oplus$ is a symbol of fuzzy addition.

$$\textit{Event time}=D\oplus TNOW \tag{4.4}$$

As a result of Equation 4.4, the event time in the event list is also a fuzzy number. Figure 4.2(a) presents an example of an event list in FDES. The simulation time advances by finding the event with the smallest event time. Because the event times are fuzzy numbers in FDES, fuzzy ranking must be performed to find the event with the smallest time in the event list. The simulation time will be updated based on the smallest event time, and new events will be added to the event list based on the simulation logic as the simulation time advances. In summary, FDES is different from DES in:

1) Storing fuzzy event times in the event list rather than crisp event times in DES
2) Using fuzzy addition for calculating the event times rather than the addition of crisp numbers in DES
3) Tracking the simulation time *TNOW,* as a fuzzy number rather than as a crisp number in DES
4) Using fuzzy ranking to find the smallest event time rather than ranking crisp numbers in DES

TNOW=tri(3.5, 4, 4.2)

Event ID	Event Time
2	tri(4.5, 6, 8)
3	tri(5.5, 7, 7.5)
6	tri(7, 8, 10)

(a)

Advancing the simulation time

TNOW=tri(4.5, 6, 8)

Event ID	Event Time
3	tri(5.5, 7, 7.5)
6	tri(7, 8, 10)

(b)

Figure 4.2 Event list in FDES (a) event list at time TNOW= tri(3.5, 4, 4.2) (b) event list at time TNOW= tri(3.5, 4, 4.2)

Fuzzy ranking produces some challenges in FDES because some of the fuzzy event times may overlap and it is possible to rank them in different orders. For example, the fuzzy times of Event 2 and Event 3 in Figure 4.2 (a) overlap; therefore, it is possible that Event 2 happens after or before Event 3 depending on the ranking method. Different choices for the ranking of these two events may produce different simulation outcomes. These outcomes are called “system evolutions”, where the chronological orders of the events in the event list will be different in each system evolution (Anglani et al. 2000).

Furthermore, the overlap of fuzzy times in FDES may produce the problem of time paradox, or in other words, going backwards in the simulation time (Perrone et al. 2001). For example, if Event 2 is selected as the next event in the event list shown in Figure 4.2(a), the simulation clock will be the time of Event 2, tri(4.5, 6,8) (Figure 4.2(b)). Afterwards, Event 3 will be fired and the simulation clock will be updated to the time of Event 3, tri(5.5, 7, 7.5). Logically, the simulation should always go forward in time and Event 3 should be greater than Event 2 for all membership degrees. However, the maximum possible time in Event 2 (8) is higher than the maximum possible value in Event 3 (7.5), so the simulation does not always adhere to the logical sequence.

Different methods are proposed in the literature for ranking fuzzy numbers in FDES. Most of these ranking methods consider one system evolution and update the simulation time to be equal to the event that is ranked the smallest. A brief summary of these methods is provided below:

- Azzaro et al. (1997) used the integral method, proposed by Liou and Wang (1992), to rank fuzzy numbers in DES; this approach is based on a value α ranging between 0 and 1 that represents the degree of optimism of the decision maker.
- Nguyen and Le (1997) used the expected existence measure (EEM) wherein the possibility that an event occurred before a specified time t is calculated.
- Perrone et al. (2001) used a fuzzy rule-based system, proposed by Klir and Yuan (1995), to rank the overlapping fuzzy numbers in the event list.
- Zhang et al. (2005) used a fuzzy ranking measure, proposed by Tran and Duckstein (2002), wherein the distance from the fuzzy set to a crisp minimum or maximum value is calculated to obtain a fuzzy distance measure. This minimum or maximum value is calculated based on the support of all fuzzy sets to be ranked at a specific time in the simulation. The fuzzy number that is closest to the minimum and furthest from the maximum is considered the smallest fuzzy number.

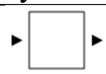
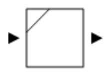
Another ranking trend in FDES attempts to consider all possible system evolutions. The ranking result is changed by randomly altering a parameter that is defined within the ranking method. The simulation model is run a number of times with different parameters that are randomly chosen in order to consider various possible system evolutions. Afterwards, the average of all evolution outcomes is considered the simulation's output. In these ranking approaches, fuzzy event times are truncated based on the lower and upper limits that are logically imposed by the ranked order of events. Anglani et al. (2000) proposed an approach based on the EEM (Nguyen and Le 1997) to rank the fuzzy numbers in FDES. First, an ordering rule among the fuzzy sets is defined. Secondly, the next event is determined by assigning crisp independent values to two algorithm parameters: the first parameter influences the event occurrence time while the second parameter affects the first event that will occur. Mehdi et al. (2005) proposed an approach similar to that of Anglani et al. (2001) but used an inverse function of the expected existence measure (EEM^{-1}). They claimed that all possible system evolutions can be generated by randomly changing the parameter assigned to EEM^{-1} between 0 and 1.

Despite the different approaches proposed for ranking fuzzy event times in FDES, there is no consensus in the literature on the best ranking approach. Zhang et al. (2005) recommended using the criteria suggested for traditional fuzzy ranking methods (such as those suggested in Chen et al. 1992) for evaluating the ranking methods in FDES. However, these criteria are not adapted for FDES and most of them are not applicable to ranking approaches based on different system evolutions. A more promising evaluation approach for ranking methods in FDES is to analyze their performance in fuzzy simulation models. For example, fuzzy simulation can be applied on simple simulation examples, so that the logic of the simulation can be easily verified by manual calculations. The result of fuzzy simulation in these situations should be compatible with the simulation logic and analytically calculated results.

4.3 IMPACT OF FUZZY RANKING METHODS ON FDES

The behaviour of fuzzy ranking methods and updating the simulation time are analyzed in this section. CYCLONE modelling symbols are used for this analysis since these symbols are very common in construction simulation; The CYCLONE simulation approach has been used extensively over the last 35 years for modelling various types of construction operations. Also, many elements of the CYCLONE methodology are common in simulation methodologies introduced after CYCLONE, such as STROBOSCOPE (Martinez 1996) and COOPS (Liu 1991). The elements of CYCLONE are listed in Table 4.1 (Halpin and Riggs 1992). NORMAL and COMBI elements represent unconstrained and constrained activities, respectively. Each COMBI must be preceded by a QUEUE. An entity cannot pass through a COMBI until all of the QUEUEs preceding a COMBI contain an entity. The duration of an activity is the main input of NORMAL and COMBI. These elements delay an entity by the specified duration of the activity. A priority is also assigned to a COMBI to indicate the priority of one COMBI over another when competing for shared resources.

Table 4.1 Elements of CYCLONE

Name	Symbol	Description
NORMAL		Unconstrained activity
COMBI		Constrained activity
QUEUE		A queueing of the resource entities/ Idle state of the resource entities
ARROW		Directional flow of the resource entities
COUNTER		A counter that counts number of passing entities
FUNCTION		A function that can consolidate or generate resource entities

A fuzzy CYCLONE simulation model is defined as a CYCLONE simulation model in which the durations of activities (NORMAL or COMBI elements) are fuzzy numbers instead of probability distributions. To process the events generated from a fuzzy CYCLONE methodology, a FDES is required. In this section, the performance of different FDES ranking approaches on different patterns of CYCLONE elements is analyzed. These patterns are introduced by Halpin and Riggs (1992) as basic patterns that can be used to simulate a variety of construction operations. These simple patterns allow us to verify the behavior of each ranking method through manual calculations.

As a simple combination, consider n sequential activities represented by sequentially connected NORMAL elements Normal(1) to Normal(n) (Figure 4.3). Assume the durations of Normal(1) to Normal(n) are fuzzy numbers D_1 to D_n respectively. In the FDES, the event related to the completion time of the Normal(i), $1<i\leq n$, is inserted in the event list when the current simulation time is equal to Normal($i-1$). This is because the element Normal($i-1$) is a predecessor of Normal(i). Therefore, the event for

completion time of Normal(i) will only be inserted in the event list after Normal($i-1$) has been completed. As a result, no ranking is required between the events that are generated for the activities in this sequence. Assuming the entity arrives at Normal(1) at fuzzy time T_1, for all ranking approaches that are based on one system evolution, the completion time of the Normal(n), C_n, will be equal to Equation 4.5, where $\oplus$ represents the addition of fuzzy numbers.

$$C_n = T_1 \oplus D_1 \oplus D_2 \oplus \ldots \oplus D_n \tag{4.5}$$

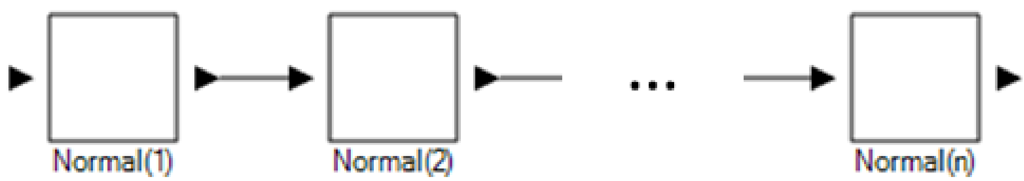


Figure 4.3 Sequentially connected activities in CYCLONE

However, if two or more sequences of activities are performed in parallel, these results will differ for the ranking approaches that are based on different system evolutions. This is because the event times will be updated based on the ranked order of the event times in the event list (Anglani et al. 2000).

Consider Example 1, where a NORMAL (or COMBI), Normal(1), is followed by two NORMAL elements, Normal(2) and Normal(3), to be processed in parallel (Figure 4.4). The durations of Normal(1), Normal(2), and Normal(3) are fuzzy numbers D_1, D_2, and D_3, respectively, and the arrival time of entity in Normal(1) is T_1. In this example, the event for the completion time of Normal(2) and Normal(3) will be inserted in the event list when the current event is equal to the completion of Normal(1) and the simulation time is equal to $T_1 \oplus D_1$. The event times for completion of Normal(2) and Normal(3) are calculated using fuzzy arithmetic in the FDES. Other ranking methods based on one system evolution may rank the completion times of Normal(2) and Normal(3) differently. These different rankings will impact the sequence in which these activities will be completed in the simulation. However, no matter which event is selected first, the simulation outcome will represent the completion time of Normal(2) as C_2 and the completion time of Normal(3) as C_3, according to equations 4.6 and 4.7.

$$C_2 = T_1 \oplus D_1 \oplus D_2 \tag{4.6}$$

$$C_3 = T_1 \oplus D_1 \oplus D_3 \tag{4.7}$$

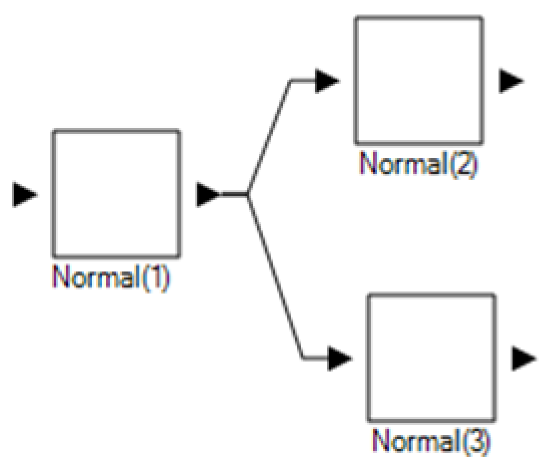


Figure 4.4 Modelling parallel activities in CYCLONE (Example 1)

On the other hand, in the approaches based on different system evolutions (Anglani et al. 2000, Mehdi et al. 2005), all different possible rankings will be considered. Assume there are two different fuzzy sets A and B for the event times C_2 and C_3 in Example 1, as illustrated in Figure 4.5(a). All ranking methods based on one system evolution will rank

A greater than *B*. However, using the ranking methods proposed in (Anglani et al. 2000, Mehdi et al. 2005), in different system evolutions, *B* will be considered greater or less than *A*. In the case of *B*>*A*, C_3 will be updated to *B'* based on the minimum possible value of *A*. The weighted average of *B* and *B'* will be calculated to find the completion time of Normal (2). This average will have a mean value greater than the mean value of *B* and a support less than that of *B* because the mean value of *B'* is greater than *B* and its support is less than *B*. Since the logical completion time of Normal (2) in Example 1 should be equal to *B*, the completion time of Normal (2) will be overestimated and its support will be underestimated using these approaches (Anglani et al. 2000, Mehdi et al. 2005).

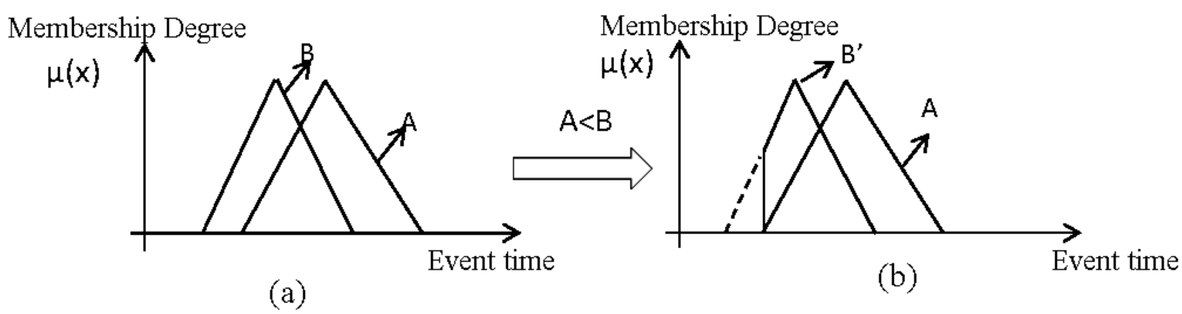


Figure 4.5 (a) Fuzzy event times A and B (b) Updating event times if A<B

The impact of different ranking approaches becomes even more significant when dealing with QUEUEs in CYCLONE models. For example, when a number of QUEUEs are followed by a COMBI, the ranking approach will impact the start and completion time of the COMBI. Assume the arrival time of the entity in each of Queue(1) to Queue(n) is T_1 and T_n, respectively, in Figure 4.6. Logically, the start time of Combi(1) is equal to max$(T_1, \ldots, T_n)$ because the COMBI cannot start until all entities arrive in Queue(1) to Queue(n).

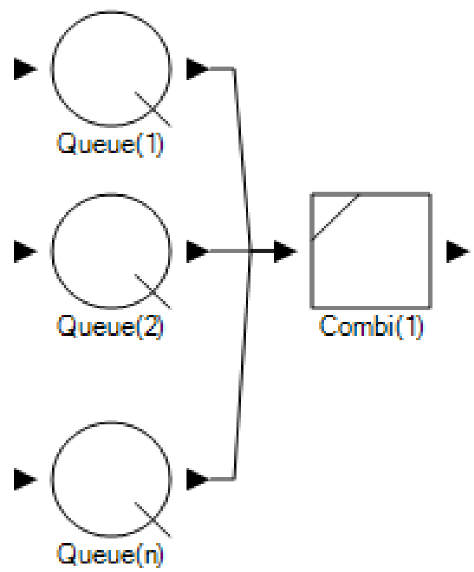


Figure 4.6 A number of QUEUEs followed by a COMBI (Example 2)

Consider Example 2, where the number of QUEUEs (n) is equal to 2. Figure 4.7(a) represents T_1 and T_2 as fuzzy numbers A and B. If A is ranked less than B in FDES, B will be the start time of Combi(1). However, if A is ranked greater than B, A will be the start of Combi(1). Logically, however, the start time of Combi(1) should be equal to max(A,B), which is greater than both A and B in this example. Therefore, in either case, the simulation time for the start time of COMBI will be underestimated. This underestimation of the simulation time is the direct consequence of the time paradox (going backward in the simulation time).

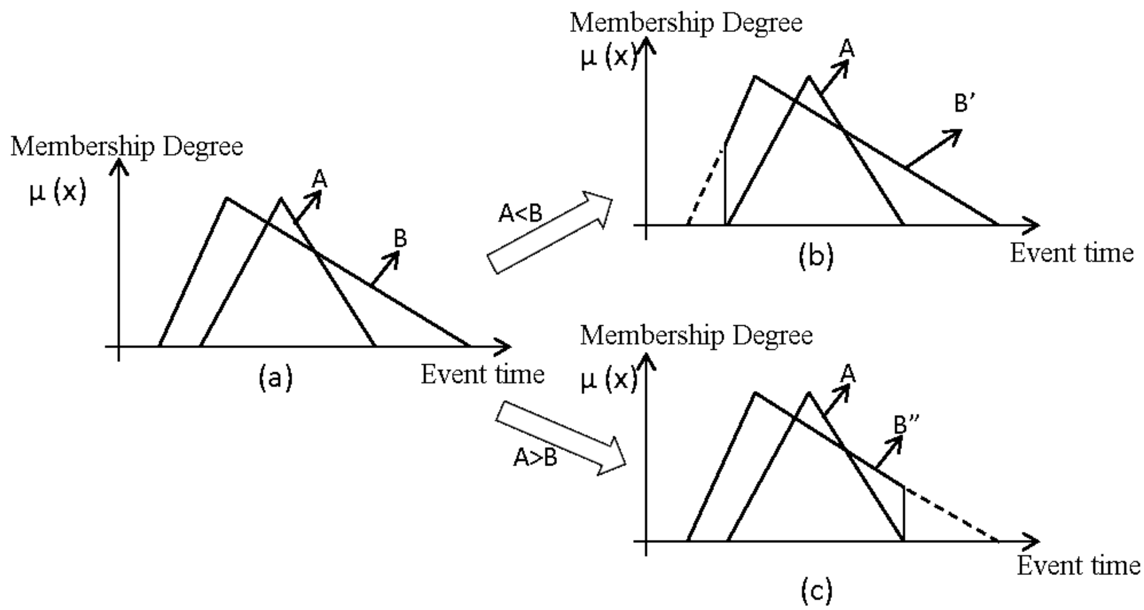


Figure 4.7 Fuzzy event times (a) Fuzzy event times A and B in FDES (b) Updating the event times when A<B (c) Updating the event times when A>B

The fuzzy ranking methods (Anglani et al. 2000, Mehdi et al. 2005) perform better in Example 2; here, both the cases $A<B$ and $A>B$ are considered and the event time of B will be updated to B' and B'' for each ranking order as indicated in Figure 4.7(b) and 4.7(c). The final time for the start of COMBI will be the weighted average of B' and A, which is greater than both A an B and can reduce the problem of underestimation of the simulation time.

A similar situation occurs when a FUNCTION element performs a consolidating operation in CYCLONE. Assume the arrival times of the entities to be consolidated are T_1 to T_n. Logically, an entity should get transferred out of the consolidated element when the simulation time equals max$(T_1,\ldots,T_n)$. Ranking methods, however, will choose one of the times T_1, $T_{2\ldots}$ or T_n to represent when the entity will be transferred out of

FUNCTION depending on ranking order. The problem in this situation is exactly the same as the situation in Example 2: none of the event times alone can represent the logical event time which should be equal to max$(T_1,\ldots,T_n)$. The approaches taken in (Anglani et al. 2000, Mehdi et al. 2005) are also more logical in this situation.

Another pattern in CYCLONE is a QUEUE followed by a number of COMBIs. Consider Example 3 where Queue(2) is followed by two COMBIs, Combi(1) and Combi(2). Combi(1) and Combi(2) have other starting conditions of Queue(1) and Queue(3), respectively (Figure 4.8). The priority of Combi(1) is higher than Combi(2) when competing for a resource. The arrival times of the entities at Queue(1), Queue(2), and Queue(3) are T_1, T_2, and T_3 respectively. Also, the duration of Combi(1) is D_1 and the duration of Combi(2) is D_2.

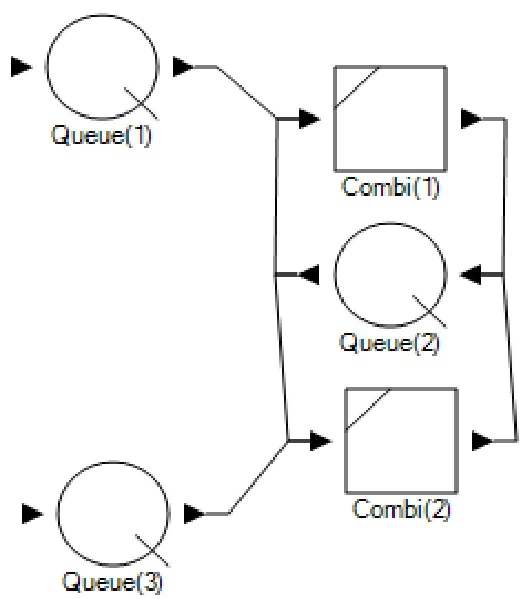


Figure 4.8 A QUEUE followed by a number of COMBIs (Example 3)

The choice of ranking method in FDES can greatly impact the start time of Combi(1) and Combi(2) in Example 3. Table 2 indicates different possible rankings of the event times and their logical impacts on the start times of the COMBIs in Example 3. When the event times overlap, different ranking methods may order the event times differently and produce different start times for Combi(1) and Combi(2). Each ranking method that considers one system evolution is only able to consider one ranking possibility. Theoretically, however, more than one case may be possible when the event times are uncertain.

Table 4.2 Impact of different orders of event times on the start times of Combi(1) and Combi(2) in Example 3

Case ID	Order of times	Start time of Combi(1)	Start time of Combi(2)
1	$T_1>T_2>T_3$	max(T_1, T_2+D_2)	T_2
2	$T_1>T_3>T_2$	max(T_1, T_3+D_2)	T_3
3	$T_2>T_1>T_3$	T_2	T_2+T_1
4	$T_2>T_3>T_1$	T_2	T_2+T_1
5	$T_3>T_1>T_2$	T_1	max(T_3,T_1+D_1)
6	$T_3>T_2>T_1$	T_2	max(T_3,T_2+D_1)

Assume T_1 and T_3 are equal to *A* and *B* as indicated in Figure 4.5(a). All of the ranking approaches based on one system evolution will rank $T_1>T_3$. However, T_1 and T_3 have a large section of overlap and therefore, there is a possibility that $T_1<T_3$. This possibility is not considered by the ranking methods based on one system evolution. To better illustrate this concept, compare this situation with a similar traditional simulation model with triangular probability distributions for the activity durations. When the probability distributions of T_1 and T_3 overlap, different samples of T_1 and T_3 will produce different orders of the event times; therefore, different possible orders will be covered through

various simulation runs. Like the probabilistic case, the fuzzy ranking approaches proposed in (Anglani et al. 2000, Mehdi et al. 2005) are able to consider all of the possible cases by randomly altering the ranking parameters in different simulation runs.

In terms of other CYCLONE elements, COUNTER does not perform any operation and will not impact the outcome of the simulation. The FUNCTION element, when performing the generate operation, will not be impacted by the ranking method, because no ranking is required since only the entity will be reproduced when arriving at the generating FUNCTION element.

Based on the above analysis, fuzzy ranking approaches based on a single system evolution will produce two issues when dealing with QUEUEs and consolidating FUNCTIONs in fuzzy CYCLONE: underestimation of the simulation time and inability to consider all system evolutions. It is very difficult to favor one ranking approach over the others in these cases, because the issue of ranking is very subjective, and for “non-questionable” cases, most methods produce identical ranking orders (Prodanovic and Simonovic 2002). Fuzzy ranking approaches based on various system evolutions try to deal with these issues. However, as discussed previously, these approaches update the simulation time of parallel activities in a way that results in two other problems: overestimation of the fuzzy simulation time and underestimation of the support of fuzzy simulation time. These shortcomings of available ranking methods in FDES need to be addressed to ensure the benefits of FDES can be derived in actual construction projects. In the next section, a novel approach is proposed to updating the simulation time in

FDES to address the discussed shortcomings of underestimation/overestimation of the simulation time.

4.4 PROPOSED FRAMEWORK FOR FDES

As previously discussed, in current methodologies of FDES, the simulation time is either overestimated or underestimated. In this section, a new methodology is proposed for FDES to eliminate the problem of underestimation/overestimation of the simulation time. This approach is based on calculating the event time according to the logical dependencies in the simulation.

In all approaches of FDES, the event time is calculated by adding *TNOW* and the delay time of the event. In the proposed methodology in this research, the event time is calculated based the maximum of all of the fuzzy times of the previous events that must first be executed before the current event. Assume that a simulation logic will impose that event e should be generated only after all of the events $e_1, e_2 \ldots e_n$ are executed. These events are referred as predecessor events for event *e*. For example, when event *e* represents the completion of an activity, events $e_1 \ldots e_n$ can represent the events for the completion of the predecessor activities or the arrivals of the required resources. Assume $T_1, T_2 \ldots T_n$ are the event times for $e_1, e_2 \ldots e_n$, respectively. After all of the predecessor events of *e* are executed, the event time for *e*, T_e is calculated as:

$$T_e = \max(T_1 \ldots T_n) \oplus D \tag{4.8}$$

In Equation 4.8, D is the delay time of event e. Also, max represents the maximum of fuzzy sets based on fuzzy arithmetic (Equations 4.2 and 4.3). For each α∈[0,1], The maximum (max) of (n) alpha-cuts, $T_{1\alpha}$... $T_{n\alpha}$ can be calculated as:

$$\max(T_{1\alpha}\ldots T_{n\alpha})=[\max(l_{1\alpha}\ldots l_{n\alpha}),\ \max(U_{1\alpha}\ldots U_{n\alpha})] \tag{4.9}$$

In Equation 4.9, $l_{i\alpha}$ is the lower bound and $U_{i\alpha}$ is the upper bound of the alpha-cut, $T_{i\alpha}$. The calculated alpha-cuts are then aggregated over all α∈[0,1] according to Equation 4.3 to calculate the final fuzzy membership function for $\max(T_1 \ldots T_n)$.

If event e represents the completion of an activity, the delay time D will be equal to the duration of that activity in Equation 4.8. In this case, $\max(T_1 \ldots T_n)$ represents the start time of the activity. The following list summarizes the steps in the proposed framework of FDES:

1) Initialize the event list and simulation time
2) For each event whose predecessor events have already been executed:
    a. Calculate the event time using Equation 4.8
    b. Insert the event to the event list
3) If there are no events in the event list, the simulation is complete, and terminate the process.
4) Otherwise, find the next event, e, by selecting the event with the smallest time from the event list. Because the event times are fuzzy numbers in FDES, fuzzy ranking is required to find the smallest event time.
5) Remove the event, e, from the event list.

6) Update the simulation time based on the time of the event, T_e ($TNow=T_e$).

7) Execute event e.

8) Proceed to step 2.

For example, consider the behaviour of the proposed approach in Example 2, in which the arrival times of the entities in Queue(1) and Queue(2) are fuzzy numbers A and B that are represented in Figure 4.7(a). The start time of Combi(1) will be calculated as max*(A,B),* because the arrival of the events in Queue(1) and Queue(2) are predecessor events for starting Combi(1). Figure 4.9 shows the resulting fuzzy set. This maximum value will improve the results of the FDES compared to other approaches that choose either A or B as the start time of the COMBI. This is because both A and B are smaller than $\widetilde{max}(A,B)$ which is the logical start time of Combi(1).

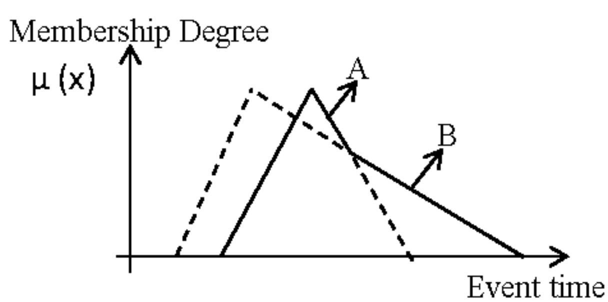


Figure 4.9 Start time of Combi(1) in Example 2 in the proposed approach

In a similar manner, for the consolidating FUNCTION, assume the arrival times of *(n)* entities to be consolidated are $T_1 \ldots T_n$. Any other approach that is based on one system evolution will choose only one of $T_1 \ldots T_n$ for the time that the entity will pass the consolidating FUNCTION. As discussed in the previous section, when these times have large overlapping sections, choosing only one of the event times will result in underestimation of the simulation time. In the proposed approach to updating the simulation time, the actual value of $max(T_1, ..., T_n)$ is calculated to represent the time that the entity is transferred out of the FUNCTION element, and hence the problem of underestimation of the simulation time is eliminated.

Furthermore, the problem of overestimation of the simulation time does not exist in the proposed approach. Changing the simulation time *TNOW* based on the events in the event list is the cause of overestimation of the simulation time in the previous FDES methodologies. In the proposed methodology, the simulation time *TNOW* is equal to the time of the event that is ranked as the smallest event in the event list. Therefore, the problem of overestimation of the simulation time will not occur.

The proposed approach to updating the simulation time when facing COMBIs and consolidating FUNCTIONs can be combined with any fuzzy ranking approach. The choice of fuzzy ranking method will not impact the results of this approach, except when facing a QUEUE that is going to a number of COMBIs, as in Figure 4.8. In the latter case, the proposed approach has the limitation of considering only one possible path of the entity depending on the fuzzy ranking method (like all FDES ranking approaches based on one system evolution). However, considering the minimal impact of the ranking method in the proposed approach, any method of ranking fuzzy numbers can be employed. For example, ranking methods such as integral method (Liou and Wang 1992) and the Chen and Chen method (Chen and Chen, 2003), which consider both the center of gravity and variance of the fuzzy set in the ranking function, are appropriate. Therefore, for increased performance of the simulation, simple ranking methods based on defuzzification, such as the centroid method (Wang et al. 2006, Yager 1980) are recommended for use in the proposed approach.

All of the discussed FDES approaches previously proposed (for ranking and updating the simulation time) either overestimate or underestimates the simulation time of FDES. The proposed approach to calculating fuzzy event times advances the state of the art of FDES by eliminating such problems. In the next section, the implementation of FDES and fuzzy CYLONE are explained.

4.5 IMPLEMENTATION

A FDES engine is developed by extending the capabilities of Simphony.NET software. Simphony.NET is a DES program for construction process modelling that employs object-oriented programming. Two main components were added to Simphony.NET in order to perform FDES (Sadeghi et al. 2013): (1) fuzzy set class, which is capable of defining a fuzzy set of any shape, taking its alpha-cuts, performing fuzzy arithmetic, and graphically representing the fuzzy number; and (2) simulation class, which is capable of scheduling an event with fuzzy time intervals and providing the current fuzzy simulation time. The simulation class uses fuzzy arithmetic to calculate the fuzzy event time and stores the fuzzy event time in an attribute of the simulation engine. Further details of this implementation are provided in Appendix B of this dissertation.

Using the implemented FDES engine, a fuzzy CYCLONE template in Simphony.NET is developed. This template contains the following elements: fuzzy COMBI, fuzzy NORMAL, fuzzy QUEUE, fuzzy FUNCTION, and RANKING. The COUNTER element that is already available in the CYCLONE template of Simphony.NET can be used in fuzzy CYCLONE as well. RANKING enables us to choose to implement one of several different ranking methods: integral (Liou and Wang 1992), centroid (Wang et al. 2006, Yager 1980), or Chen and Chen (2003). An attribute is also defined to choose between the traditional method of calculating the event times in FDES and the proposed approach to updating the simulation time in FDES. With the graphical user interface, users can develop simulation models by dragging and dropping elements into the simulation environment. A snapshot of the simulation environment is represented in Figure 4.10.

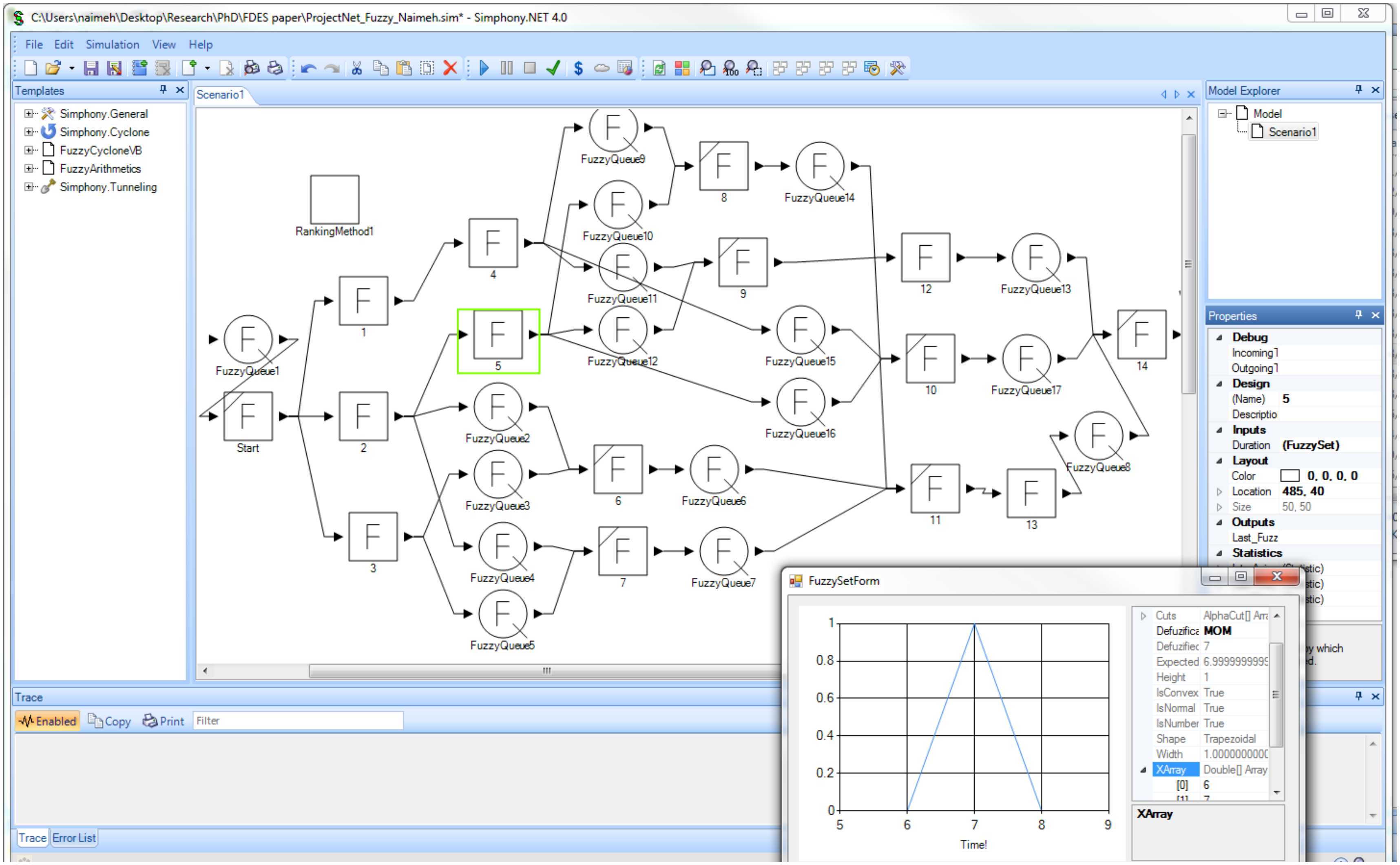


Figure 4.10 Project network modelled fuzzy CYCLONE template

A practical example and an actual case study of construction projects are provided in the following sections to illustrate the validity of the proposed FDES methodology and its potential applications of FDES in construction management.

4.6 ILLUSTRATIVE EXAMPLE OF A BUILDING CONSTRUCTION PROJECT NETWORK

The following example of a project network illustrates the performance of the proposed approach to calculate the event times in FDES. To compare the results with a benchmark that can be calculated analytically, a construction project network of a building construction project was adapted from (Halpin and Riggs 1992). This simple project network is chosen for this example to be able to analytically calculate and verify the FDES results. The completion time of a project network with fuzzy activity durations can be analytically calculated similarly to traditional CPM project network calculations but by performing fuzzy addition and fuzzy maximum operations (Lorterapong and Moselhi 1996, Prade 1979). The project data are shown in Table 3. The duration of the activities are defined as either constant or triangular fuzzy numbers. The fuzzy numbers represent the uncertainty of the activity durations.

Table 4.3 An example list of the activities, durations, and logical dependencies of building construction operation (adapted from Halpin and Riggs 1992)

ID	Activity	Duration	Predecessor
1	Prefab Wall Forms	constant(2)	-
2	Excavate Cols and Walls	constant(3)	-
3	Let Elec and Mech Subcontract	tri(3,4,8)	-
4	Deliver Wall Forms	constant(4)	1
5	Forms, Pour and Cure Wall & Column	tri(6,7,8)	2
6	Rough-In Plumbing	tri(5,7,10)	2,3
7	Install Conduit	tri(9,11,15)	2,3
8	Erect Wall Forms and Steel	constant(9)	4,5
9	Fabricate and Set Interior Column Forms	constant (6)	4,5
10	Erect Temporary Roof	tri(12,16,18)	4,5
11	Pour, Cure and Strip Walls	tri(10)	6,7, 8
12	Pour, Cure and Strip Int. Walls	tri(6)	9
13	Backfill for Slab on Grade	constant(1)	11
14	Grade and Pour Floor Slab	constant(5)	12,13

Figure 4.10 represents the simulated project network using fuzzy CYCLONE, implemented based on the FDES. When using fuzzy CYCLONE model with any of the implemented ranking methods, without applying the approach proposed in this chapter for calculating the event times, the resulting fuzzy project completion time equals tri(34, 35, 36). In this case, the maximum possible value for the project completion time with FDES is calculated as 36. When applying the proposed approach for calculating the event times in fuzzy CYCLONE, the fuzzy project completion time is calculated as shown in Figure 4.11, where the minimum value is 34 and the maximum value is 39. If this fuzzy project network is solved analytically (Prade 1979), the results are exactly the same as the results obtained using the proposed approach of calculating the event times

(Figure 4.11), confirming the fact that the proposed approach of calculating the event times is correct and matches the analytically calculated results. The results represent the uncertain range of the output with different degrees of possibility. These results may be analyzed further by obtaining the alpha-cuts of the output fuzzy set based on the desired confidence level (Mauris et al. 2001). For a confidence level equal to λ, the alpha-cut at a level of α equal to 1-λ will represent this range. For example, for a confidence level λ equal to 0.8, the project will be completed within 34.2 to 37.4 days. This range is obtained by getting the alpha-cut at the level of α=0.2, as illustrated in Figure 4.11.

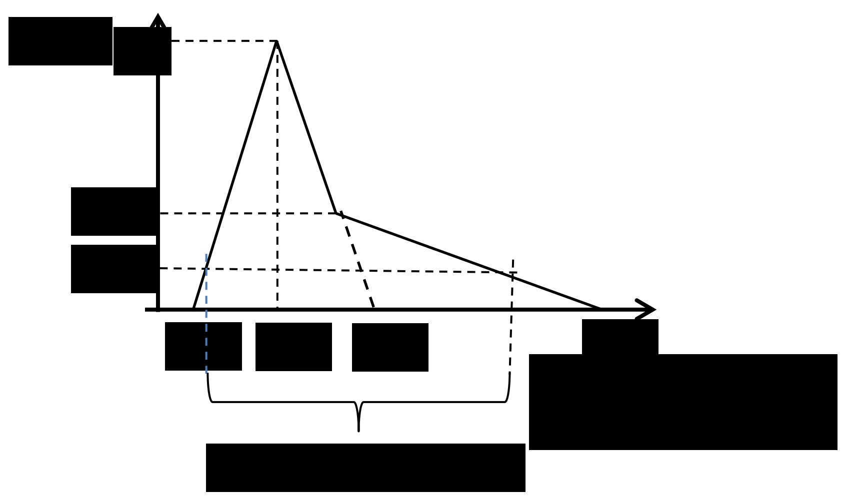


Figure 4.11 Fuzzy project completion time calculated using fuzzy CYCLONE approach

The case study in the next section further illustrates the practical benefits of FDES.

4.7 CASE STUDY OF TUNNELING OPERATION

A case study of the City of Edmonton NEST tunnel (Chung et al. 2006) is used to show the benefits of using FDES to represent the uncertainty of TBM (Tunnel Boring

Machine) penetration rate in the simulation model of a tunneling operation. Tunneling projects include three major operations: excavation, dirt removal, and lining. TBM penetration rate is one of the most important factors impacting the production rate of tunneling operations (Chung et al. 2006). Shaheen et al. (2009) developed a fuzzy expert system to model various factors that impact the TBM penetration rate. They generated the rules' antecedents and consequents using knowledge gained from interviewing tunneling experts. They later employed their developed fuzzy expert system in the NEST tunnel to study the effect of modelled factors on the final simulation output. In the case study, the tunnel is divided into 9 different segments based on the inputs of the fuzzy expert system. A different TBM rate is estimated for each of the soil segments in the simulation model. The output of the fuzzy expert system is defuzzified using the centre of area method (CoA) to represent the TBM advance rate as a crisp number.

The case study of the City of Edmonton NEST tunnel is recreated in the developed FDES program based on the inputs provided by Chung et al. (2006) and Shaheen et al. (2009). Shaheen et al. (2009) used the defuzzified outputs of a fuzzy expert system, which are crisp numbers, as the inputs of the simulation model; however; representing the estimated penetration rate as a crisp number disregards the uncertainty that exists in its estimated value.

> The if-then rules in fuzzy expert systems are defined with linguistic terms; for example a rule can be, If TBM age is *low* and operator's experience is *high*, then the TBM penetration rate is *high*.

A fuzzy set is defined for each linguistic term of a variable, such as *low*, *medium* and *high* temperature. The fuzzy sets defined on the output of a fuzzy expert can represent the uncertainty of the value estimated by the fuzzy expert system (Janssen et al. 2010). Therefore, some researchers propose to represent the output of a fuzzy expert system as a fuzzy number that matches a linguistic term such as *high* or *low*, instead of defuzzifying the output and representing it as a crisp number. In order to determine the linguistic term for the output of a fuzzy expert system, the Euclidean distance from the output of the fuzzy expert system to the membership functions of the outputs' linguistic terms is calculated; the shortest Euclidean distance to a linguistic term determines the best linguistic term to characterize the output of the fuzzy expert system (Fayek and Oduba 2005).This approach is used in the tunneling case study to represent the output of the fuzzy expert system, which predicts the TBM penetration rate. The penetration rate of each segment will match one of the *low*, *medium*, or *high* linguistic terms. The fuzzy numbers of the output linguistic terms of the fuzzy expert system for the tunneling penetration rate are represented in Figure 4.12. The fuzzy numbers of the penetration rate calculated for each segment are then used directly as the inputs to the FDES model. For probabilistic inputs of this case study (such as traveling times, time between break-downs of TBM, time to repair TBM, etc.), the mean values of the probability distributions of other simulation inputs provided by Chung et al. (2006) and Shaheen et al. (2009) are used. This is because the FDES program is not at this point capable of considering both fuzzy and probabilistic inputs. In Chapter 6, a framework that can handle both fuzzy and probabilistic distributions will be developed.

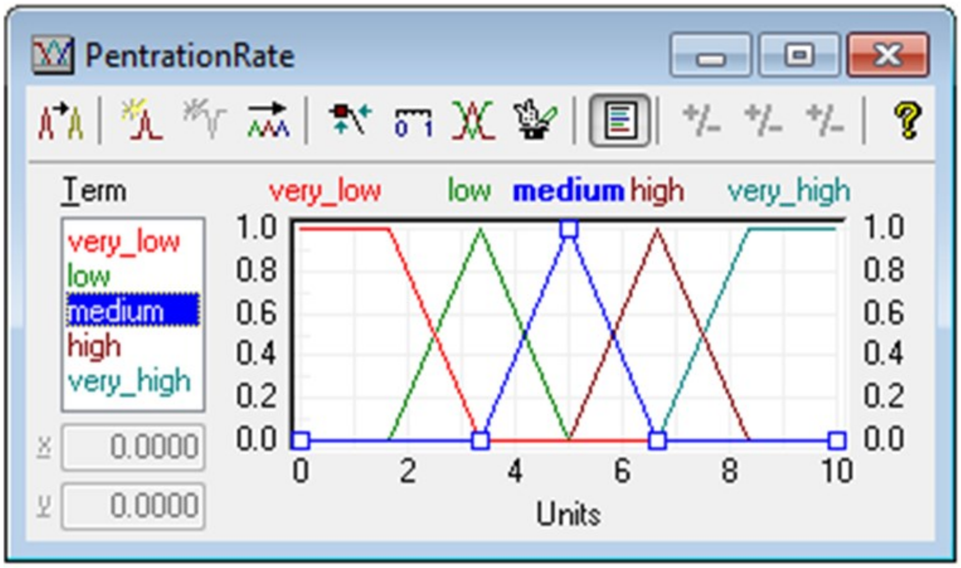


Figure 4.12 Membership functions for linguistic terms of penetration rate of TBM machine (Adapted from Shaheen et al. 2009)

The actual average construction production rate for the NEST tunnel based on the field data was reported equal to 8.87 meters/shift Chung et al. (2006). A production rate of 9.75 meters/shift (averaging 10 simulation runs) is reported by Shaheen et al. (2009) when using the defuzzified values of fuzzy expert system for penetration rate. The estimated production rate by Shaheen et al. (2009) has about 10% error compared to the actual production rate. All of the implemented approaches of previous FDES frameworks produce the same result for the production rate of tunneling that is presented in Figure 4.13. On the other hand, the membership function of the production rate estimated by the proposed FDES framework is smaller compared with previous FDES frameworks as presented in Figure 4.14. This is because previous FDES frameworks that are implemented underestimate the project completion time resulting in overestimation of production rate. The membership function of the production rate estimated by both proposed and previous FDES frameworks has a full membership degree of 9.65 meters/shift. However, the supports of these two fuzzy numbers are different. The minimum and maximum values of the support of the proposed FDES framework are 8.21 and 10.53 meters/shift, respectively, which is smaller compared with the minimum and maximum of the support of previous FDES frameworks, 8.28 and 10.63 meters/shift. Furthermore, the defuzzified value of the production rate using the centroid defuzzification method is 9.40 meters/shift in the proposed FDES framework (Figure 4.13), which is closer to the actual production rate of tunneling operation compared with the defuzzified value of the previous FDES frameworks which is 9.46 meters/shift (Figure 4.14). Therefore, in this case study, the defuzzified value of the production rate of all FDES frameworks are closer to the actual production rate of tunnelling operation compared with DES, while the proposed FDES produce the closest results to the actual production rate. In addition to providing a more accurate defuzzified value, the actual benefit of FDES compared with DES is in its capability to model subjective uncertainty. The provided range using FDES provides a more realistic understanding of the imprecision of the simulation output. Thus, it allows the simulation analyst to understand the real impact of subjectivity on the simulation outcome that is due to the use of linguistic expression and expert knowledge in estimating simulation inputs.

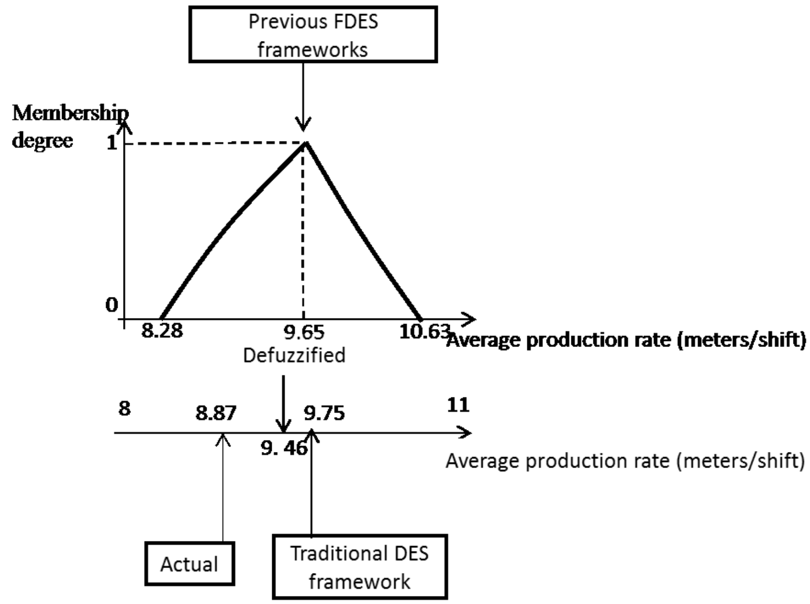


Figure 4.13 . Comparison of average tunneling production rate calculated using previous FDES frameworks that are implemented with DES, and actual tunneling production rate

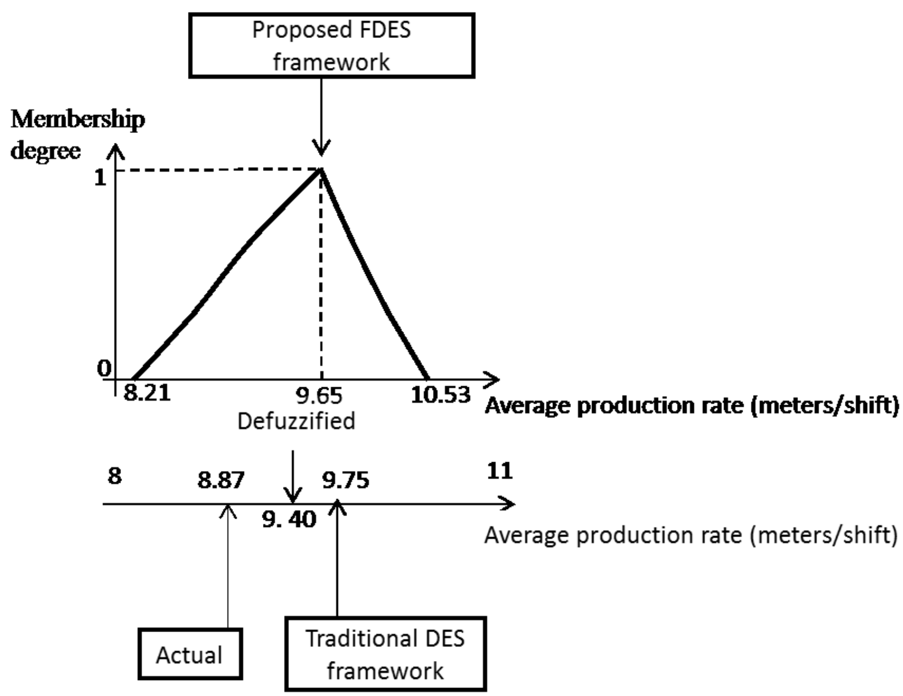


Figure 4.14 Comparison of average tunneling production rate calculated using the proposed FDES framework with DES and actual tunneling production rate

4.8 CONCLUDING REMARKS

Integrating fuzzy logic with discrete event simulation (DES) of construction projects provides significant advantages in modelling uncertainty resulting from linguistically-expressed expert knowledge. However, the available approaches of fuzzy discrete event simulation (FDES) have major shortcomings in overestimation/underestimation of the simulation time as illustrated in this chapter. A new approach for calculating the event times in FDES to eliminate these shortcomings was proposed. The proposed approach was tested with different structures of CYCLONE modelling elements and was shown to eliminate the problem of overestimation/underestimation of the simulation time.

Additionally, the proposed approach of FDES was implemented and its performance validated by comparing its results to the analytically-calculated results of a practical example of a project network. The results of the proposed FDES methodology produce the same results as the analytically-calculated results, while all other implemented approaches of FDES underestimate the project completion time. Furthermore, an actual case study of a tunnelling construction operation was presented.

The penetration rates of the TBM machine are defined as fuzzy numbers to present the subjective uncertainty due to the use of linguistic terms. The productivity of the FDES is also presented as a fuzzy number that includes the actual productivity of tunneling operation. The defuzzified value of the output of the FDES is closer to the actual productivity in the case study compared with DES and previous FDES frameworks. However, the true advantage of FDES is in its capability of representing the subjective uncertainty that was not possible to consider in DES. The representation of subjective uncertainty allows the simulation analyzer to comprehend the imprecision of the simulation outcome resulted due to the use of linguistic expression and expert judgement.

The proposed FDES approach is only capable of calculating the fuzzy simulation time and productivity as the simulation output. Approaches for calculating performance measures such as waiting time, utilization, and queue length are provided, for FDES, in the next section.

4.9 REFERENCES

Ahuja, H. N., & Nandakumar, V. (1985). Simulation model to forecast project completion time. Journal of construction engineering and management, 111(4), 325-342.

Anglani, A., Grieco, A., Nucci, F., Semeraro, Q., and Tolio, T. (2000). A new algorithm to rank temporal fuzzy sets in fuzzy discrete event simulation. Proc., Fuzzy Syst. Conf., IEEE, San Antonio, TX, 2, 923–928.

Ayyub, B. M. , and Haldar, A. (1984). Project scheduling using fuzzy set concepts. J. Constr. Eng. Manage. , 110 (2), 189–204.

Azzaro, C., Floquet, P., Pibouleau, L. and Domenech, S. (1997). A fuzzy approach for performance modeling in batch plant: Application to semiconductor manufacturing. IEEE Trans. Fuzzy Syst., 5(3), 338–357.

Chen, S. J. J., Hwang, C. L., Beckmann, M. J., and Krelle, W. (1992). Fuzzy multiple attribute decision making: methods and applications. Springer-Verlag, New York, NY.

Chen, S. J., and Chen, S. M. (2003). A new method for handling multicriteria fuzzy decision-making problems using FN-IOWA operators. Cybern. Syst., 34(2), 109–137.

Chung, T. H., Mohamed, Y., & AbouRizk, S. (2006). Bayesian updating application into simulation in the North Edmonton Sanitary Trunk tunnel project. J. Cons. Eng. Manage., 132(8), 882-894.

Dubois, D. J. (1980). Fuzzy sets and systems: theory and applications. Academic Press, New York, NY, 144.

Fayek, A. R., & Oduba, A. (2005). Predicting industrial construction labor productivity using fuzzy expert systems. Journal of construction engineering and management, 131(8), 938-941.

Guyonnet, D., Bourgine, B., Dubois, D., Fargier, H., Côme, B., & Chilès, J. P. (2003). Hybrid approach for addressing uncertainty in risk assessments. Journal of Environmental Engineering, 129(1), 68-78.

Hajjar, D., and AbouRizk, S. (1999). Simphony: an environment for building special purpose construction simulation tools. Proc. Winter Simul. Conf., Phoenix, AZ, 2, 998–1006.

Halpin, D. W. and Riggs L.S. (1992). Planning and analysis of construction operations. John Wiley & Sons, Hoboken, NJ.

Janssen, J. A. E. B., Krol, M. S., Schielen, R. M. J., Hoekstra, A. Y., & de Kok, J. L. (2010). Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models. Ecological Modelling, 221(9), 1245-1251.

Klir, G. J., and Yuan, B. (1995). Fuzzy sets and fuzzy logic. Prentice Hall, Upper Saddle River, New Jersey.

Lorterapong, P., & Moselhi, O. (1996). Project-network analysis using fuzzy sets theory. Journal of Construction Engineering and Management, 122(4), 308-318.

Liou, T. and Wang, M. (1992). Ranking fuzzy numbers with integral value. Fuzzy Sets Syst., 50(3), 247–255.

Liu, L. Y. (1991). COOPS: Construction object-oriented simulation system. Doctoral dissertation, Univ. Michigan, Ann Arbor, MI.

Martinez, J. C. (1996). STROBOSCOPE: State and resource based simulation of construction processes. Doctoral dissertation, University of Michigan, Anne Arbor, MI.

Martinez, J. C., and Ioannou, P. G. (1999). General-purpose systems for effective construction simulation. J. Constr. Eng. Manage., 125(4), 265–276.

Mauris, G., Lasserre, V., & Foulloy, L. (2001). A fuzzy approach for the expression of uncertainty in measurement. Measurement, 29(3), 165-177.

Mehdi, R., Araar, A., and Khali, H. (2005). New fuzzy ranking algorithm for discrete event simulation. IEEE Proc. Electron. Circuits Syst., 1–4.

Nguyen, Q., & Le, T. (1997). A fuzzy discrete-event simulation model. Proceedings of Australia-Pacific Forum on Intelligent Processing and manufacturing of Materials, Gold Coast, Queenland, Australia, July 14-17 (Vol. 7).

Perrone, G., Zinno, A., and La Diega, N. (2001). Fuzzy discrete event simulation: A new tool for rapid analysis of production systems under vague information. J. Intell. Manuf., 12(3), 309–326.

Prade, H. (1979). Using fuzzy set theory in a scheduling problem: a case study.Fuzzy sets and systems, 2(2), 153-165.

Prodanovic P., and Simonovic S. P. (2002). Comparison of fuzzy set ranking methods for implementation in water resources decision-making. Can. J. Civ. Eng. 29, 692–701

Sadeghi, N., Fayek, A. Robinson, and Mosayebi, S. P. (2013). Developing a fuzzy discrete event simulation framework within a traditional simulation engine. Proc. Int. Fuzzy Syst. Assoc. Congress, Edmonton, Canada, 1102–1106.

Shaheen, A. A., Fayek, A. Robinson, & AbouRizk, S. M. (2009). Methodology for integrating fuzzy expert systems and discrete event simulation in construction engineering. Can. J. Civil Eng., 36(9), 1478-1490.

Tran, L., and Duckstein, L. (2002). Comparison of fuzzy numbers using a fuzzy distance measure. Fuzzy Sets and Syst., 130(3), 331–341.

Wang, Y. M., Yang, J. B., Xu, D. L., and Chin, K. S. (2006). On the centroids of fuzzy numbers. Fuzzy Sets Syst., 157(7), 919–926.

Yager, R. R. (1980). On a general class of fuzzy connectives. Fuzzy Sets and Syst., 4(3), 235–242.

Zadeh, L. A. (2002). Toward a perception-based theory of probabilistic reasoning. Rough sets and current trends in computing, L. Polkowski, A. Skowron, eds., Springer-Verlag, Berlin, Germany, 46–48.

Zhang, H., Tam, C. M., & Li, H. (2005). Modeling uncertain activity duration by fuzzy number and discrete-event simulation. European journal of operational research, 164(3), 715-729.

CHAPTER 5 ANALYSIS OF QUEUES IN FUZZY DISCRETE EVENT SIMULATION[2]

5.1 INTRODUCTION

FDES enables the consideration of subjective uncertainty in construction simulation, which was not possible in the traditional DES framework. However, current FDES frameworks are only capable of calculating the simulation time and project completion time; methodologies for calculating other performance measures such as average queue length and waiting time have not yet been developed. Queue performance measures such as average queue length and waiting time are important in finding bottlenecks and optimizing the number of resources for construction projects (Halpin and Riggs 1992). Many applications of construction projects analyze average waiting time as an important output of simulation models (Lu 2003, Martínez 1998, Song and Eldin 2012, Song et al. 2008, Zeng et al. 2014). Thus, the lack of a methodology for analyzing queues confines the practicality of FDES in many construction projects.

The calculation of average queue length and waiting time in FDES is challenging because the event times in FDES are fuzzy numbers, and fuzzy arithmetic is required to calculate average queue length and waiting time. At the same time, these fuzzy event times are correlated and their correlations have to be considered when performing fuzzy arithmetic. The objective of this chapter is to provide an extension to FDES that allows calcuating average queue length and waiting time. The proposed extension to FDES will

[2] Parts of this chapter is submitted to the journal of Automation in Construction.

allow the consideration of subjective uncertainty in the analysis of queues in event-based simulation models, which broadens the opportunity for simulation-based analysis of construction projects when facing subjective uncertainty, for example due to the use of expert judgment in estimating activity durations.

This Chapter is organized as follows: Section 5.2 explains how the queues can be modelled in FDES; In Section 5.3, the proposed approach for calculating the average fuzzy queue length and waiting time in FDES is provided; In Section 5.4, the results of the developed approach are validated through analytically solved examples of simple fuzzy queueing systems; In Section 5.5, the practicality of the developed approach is illustrated through an example of an asphalt paving operation; Finally, the conclusions are discussed in Section 5.6.

5.2 QUEUES IN FDES

FDES is an event-based simulation approach and thus models a queueing systm in terms of events. As a simple example, consider a system where customers are waiting in a queue to get served by one teller. The durations for service times and/or inter-arrival times of customers are assumed to be fuzzy numbers. The events in such a system can be defined as arrival (representing the arrival of the customers to the queue) and departure (representing departure of customers from the system**)**. When a customer (customer1) arrives in a queue (i.e., the event *customer1_arrival* is executed) and the status of the teller is busy, the customer waits in the queue. On the other hand, when *customer1_arrival* is executed and a teller is available, the customer can start being serviced. In this case, the status of the teller is changed to busy and, since *customer1_arrival* has been executed, the event *customer1_departure* is inserted in the event list. The event time for customer departure will be calculated by the addition of the *TNOW* and service time. When the event of *customer1-departure* is executed, the status of the teller is changed to idle. If customer2 is waiting in the queue at this point, that customer can start service right away.

In the above queueing example, assume the event *customer1_departure* is scheduled at fuzzy time T_1 and event *customer2_arrival* is scheduled at fuzzy time T_2 (Figure 5.1(a)). If T_1 is ranked greater than T_2, the *customer1_departure* will be executed first. On the other hand, if T_1 is ranked less than T_2, the event *customer2_arrival* will be executed first. Upon this arrival, the teller is busy serving customer1 and therefore customer2 waits in the queue until the departure of customer1. In both ranking orders, the start time of service for customer2 is calculated based on the maximum of *T1* and *T2* (Figure 5.1(b)) because the start of service for customer 2 logically depends on the execution of both *customer1_departure* and *customer2_arrival*.

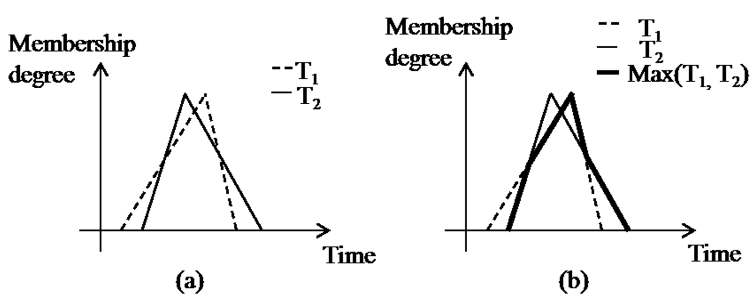


Figure 5.1 Event times in FDE simulation: (a) event times T1 and T2; (b) maximum of event times T1 and T2

As discussed in this section, FDES is able to simulate queues similar to DES. The difference is, however, that the simulation times at which an entity enters and leaves the queue are fuzzy numbers in FDES. The next section discusses the calculation of performance measures of queues in FDES.

5.3 A METHODOLOGY FOR ANALYZING AVERAGE QUEUE LENGTH AND WAITING TIME IN FDES

Average queue length and waiting time are the performance measures that are most commonly calculated for analyzing queueing systems in construction projects. The literature lacks a methodology for calculating these performance measures in FDES. This section provides a novel approach for calculating average queue length and waiting time in FDES. First, the calculation of the waiting time of an entity in FDES is illustrated. Later, the equations for deriving average queue length and waiting time based on the waiting times of individual entities are presented.

In FDES, the waiting time of an entity, W_{entity}, can be calculated by fuzzy subtraction of the time at which the entity arrived in the queue, T_a, from the time at which the entity started to be served, T_s. In subtracting the fuzzy number T_a from the fuzzy number T_s, however, the correlation between the operands (fuzzy numbers) should be considered. The importance of considering correlations of operands in fuzzy arithmetic is mentioned by Carlsson and Fullér (2004). Ignoring this correlation will result in the overestimation of the support of the calculated waiting time. For example, assume there is a special case where an entity requests a resource at fuzzy time T_a with a trapezoidal fuzzy number with parameters $a \leq b \leq c \leq d$, trap(a,b,c,d). Also, assume this entity captures that resource right away, such that the service time for the entity is also T_s=trap(a,b,c,d). In this case, by performing fuzzy arithmetic ($T_s \ominus T_a$), the waiting time of the entity will be equal to trap($a-d,b-c,c-b,d-a$). However, logically, this time should be equal to zero since the resource was available at the time that it was requested and the entity did not wait in the queue. This error is due to the correlation between T_a and T_s. Therefore, in order to correctly calculate the waiting time of entities in FDES, considering the correlation of fuzzy numbers is necessary when subtracting the event time T_a from T_s. On the other hand, in traditional DES the event times are samples from probability distributions and thus are crisp values. Therefore, normal arithmetic can be performed in DES and the problem of correlation does not occur. In the next section, a new approach is proposed for subtracting the event times in FDES that is able to consider the correlation of fuzzy event times for subtraction operation.

5.3.1 Subtracting the Event Times in FDES

As discussed in Chapter 4, event times in FDES are calculated based on one of two fuzzy operations: 1) addition, which results from adding *TNOW* and the delay time of the event; and 2) maximum, which results from the logical dependencies of the start of one event on other events in the simulation. Two event times may have correlations due to similarity of the event times from which they are derived. This correlation will impact the results of the subtraction operation that is required for calculating the waiting time of an entity. For example, assume that at a given simulation time, *TNOW*, *Event1* and *Event2* are scheduled with fuzzy time intervals TI_1 and TI_2. The event times for these events, T_{Event1} and T_{Event2}, can be calculated by performing fuzzy addition (Equation 5.1).

$$T_{Event1} = TNOW \oplus TI_1$$

$$T_{Event2} = TNOW \oplus TI_2 \tag{5.1}$$

$TNOW$ is called the parent event time of events T_{Event1} and T_{Event2}. When performing the fuzzy subtraction operation on these two fuzzy event times, the result of the subtraction of the common parts of the two events should be assumed to be 0. This is because the common section $TNOW$ in T_{Event1} is 100% positively correlated and equal to $TNOW$ in T_{Event2}. The subtraction of completely (100%) positively correlated and equal fuzzy numbers is 0 (Carlsson and Fullér 2004). Therefore, the fuzzy subtraction can be calculated according to Equation 5.2.

$$T_{Event1} \ominus T_{Event2} = (TNOW \oplus TI_2) \ominus (TNOW \oplus TI_1) =$$

$$(TNOW \ominus TNOW) \oplus (TI_2 \ominus TI_1) = (TI_2 \ominus TI_1) \tag{5.2}$$

It is possible that two events may not be correlated due to equality of their parent events, but their parents may be correlated to each other. Assume T_{Event1} and T_{Event2} are defined as Equation 5.3.

$$\begin{aligned} T_{Event1} &= T_{Event1.parent} \oplus TI_1 \\ T_{Event2} &= T_{Event2.parent} \oplus TI_2 \end{aligned} \tag{5.3}$$

When subtracting event T_{Event1} from T_{Event2}, the possible correlations of their parents should be considered:

$$T_{Event1} \ominus T_{Event2} = T_{Event1.parent} \ominus T_{Event2.parent} + TI_1 \ominus TI_2 \tag{5.4}$$

In FDES, an event time may be equal to the maximum of other events due to the logical dependencies of the start of an event on the execution of other events, as discussed in Chapter 4. For example, assume T_{Event1}=max($T_1, T_2 \ldots T_n$). When subtracting T_{Event2} from

T_{Event1}, each of the components $T_1, T_2 \dots T_n$ should be subtracted separately. The maximum of each subtraction should then be calculated to find the final results of subtraction:

$$T_{Event1} \ominus T_{Event2} = \max(T_1, T_2 \dots T_n) - T_{Event2} =$$

$$\max(T_1 \ominus T_{Event2}, T_2 \ominus T_{Event2} \dots T_n \ominus T_{Event2}) \tag{5.5}$$

On the other hand, when $T_{Event2}=\max(T_1, T_2 \dots T_n)$, the subtraction can be calculated as:

$$T_{Event1} \ominus T_{Event2} = T_{Event1} \ominus \max(T_1, T_2 \dots T_n) =$$

$$\min(T_{Event1} \ominus T_1, T_{Event1} \ominus T_2 \dots T_{Event1} \ominus T_n) \tag{5.6}$$

In Equations 5.5 and 5.6, maximum (max) and minimum (min) of fuzzy numbers is calculated based on fuzzy arithmetic. To keep track of the common components of the event times, a correlation network of the event times in FDES is developed. For event e, if event time is calculated from the time of event e_1, the parent of e is set equal to e_1 in the correlation network. The parent type is set equal to addition. On the other hand, if T_{e1} is calculated by getting the maximum from the time of events, $e_2 \dots e_n$, then the parents of event e_1 in the correlation network will be equal to $e_2 \dots e_n$ and the parent type will be equal to maximum. Figure 5.2 illustrates the correlation network built based on this example.

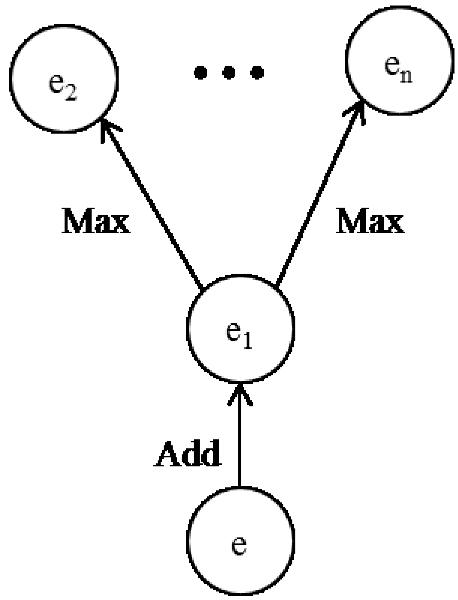


Figure 5.2 Correlation network of event times in FDE simulation

Using the proposed correlation network, the following recursive function is used to calculate the subtraction of Te_2 from Te_1 in FDES. In this pseudo code, T_{ei} represents time of event e_i, Subtract(T_1, T_2) represents $T_1 \ominus T_2$, and *e.parent* represents the parent event(s) of event *e*.

1. If e_1=e_2: Return 0
2. If the operation type of e_1 and e_2 is addition:
   a. TI_1=time interval of event e_1 ($TI_1 \oplus T_{e1.parent} = T_{e1}$)
   b. TI_2=time interval of event e_2 ($TI_2 \oplus T_{e2.parent} = T_{e2}$)
3. If e_1 is equal to one ancestor of e_2: Return Subtract(T_{e1}, $T_{e2.parent}$)$\ominus TI_2$;
4. If e_2 is equal to one ancestor of e_1: Return Subtract($T_{e1.parent}$, T_{e2})$\oplus TI_1$;
5. Else: Return Subtract($T_{e1.parent}$, $T_{e2.Parent}$)$\oplus TI_1 \ominus TI_2$;

6. If operation type of e_1 is maximum and it has n parents, *parent_1* to *parent_n*: Return Max(Subtract($T_{e1.parent\_1}$, T_{e2}), …, Subtract($T_{e1.parent\_n}$, T_{e2}))

7. If operation type of e_2 is maximum and it has n parents *parent_1* to *parent_n*: Return Min(Subtract(T_{e1}, $T_{e2.parent\_1}$), …, Subtract(T_{e1}, $T_{e2.parent\_n}$))

The above recursive procedure is able to subtract two event times in FDES by considering the correlations of the event times. Thus, this recursive procedure is used to calculate the fuzzy performance measures in FDES, as discussed in the next section.

5.3.2 Calculating Average Queue Length and Waiting Time

The proposed subtraction procedure based on the correlation network is used to subtract the start of service from the arrival time of each entity in the FDES to find an entity's waiting time in a queue. The waiting times of different entities can be used to estimate the average fuzzy queue length and waiting time in the FDES. Assume that during the simulation, entities 1 to n enter and leave a queue behind a server and their waiting time in the system is $W_1 \ldots W_n$. The average waiting time for that queue can be calculated by averaging from $W_1 \ldots W_n$ using fuzzy arithmetic. If n is large enough, this average value will be close to the steady state waiting time, W_q(Equation 5.7).

$$W_q = \lim_{n\to\infty} \left(\sum_{i=1}^{n} W_i\right) \oslash n \tag{5.7}$$

In Equation 5.7, $\oslash$ represents the division of fuzzy numbers based on fuzzy arithmetic. For calculating the average time a unit spends in the system, $\overline{W}$, the service time of each entity i, S_i, will be also added to W_i (Equation 5.8).

$$\overline{W} = \sum_{i=1}^{n}(W_i \oplus S_i)) \oslash n \tag{5.8}$$

The average length of waiting line L_q can be calculated based on the sum of waiting of all of the entities in the waiting line over the total simulation time, *TotalTime*; when n is large; this average value is close to the steady state results (Equation 5.9).

$$L_q = \lim_{n\to\infty}(\sum_{i=1}^{n}(W_i)) \oslash TotalTime \tag{5.9}$$

The average number of entities in the system, L, can be calculated based on the sum of waiting plus service time over *TotalTime* (Equation 5.10).

$$L = \lim_{n\to\infty}(\sum_{i=1}^{n}(W_i \oplus S_i) \ \oslash TotalTime \tag{5.10}$$

In equations 5.7 to 5.10, the arguments W_i, S_i and *TotalTime* are fuzzy numbers and therefore, fuzzy arithmetic must be performed in each operation. In equations 5.9 and 5.10, W_i and S_i are correlated with the *TotalTime* of simulation. However, assuming that n is large, the correlation between W_i and S_i to the *TotalTime* can be assumed to be very small and therefore can be ignored. The proposed method is implemented and validated through analytically solved examples, as discussed in the following section.

The average queue length can be also calculated based on average waiting time according to the theory proposed by John little (Little's Law) (1961):

> "The long term average number of customers in a stable system, L , is equal to the average arrival rate, λ, multiplied by the average time a customer spends in a system, W (Equation 5.11)"

$$L = \lambda * W \tag{5.11}$$

In fuzzy queueing systems, the average arrival rate, λ, can be calculated based on the total number of entities, n, over the total simulation time, $Total\_Time$. Because the simulation time is a fuzzy number, fuzzy arithmetic is required to calculate λ according to Equation 5.12.

$$\lambda = \frac{n}{Total\_Time} \tag{5.12}$$

Thus, Equation 5.10 is the same as the Little's Law when replacing, λ and W in Equation 5.11 with equations 5.10 and 5.8, respectively (Equation 5.13).

$$L = \lambda * W = \frac{n}{Total\ Time} * \sum_{i=1}^{n} (W_i \oplus S_i)) \oslash n =$$

$$\lim_{n\to\infty} (\textstyle\sum_{i=1}^{n} (W_i \oplus S_i) \ \oslash TotalTime \tag{5.13}$$

5.4 IMPLEMENTATION AND VALIDATION

The proposed approach of calculating average queue length and waiting time in FDES is implemented within the previously developed FDES engine and fuzzy CYCLONE template (as discussed in Chapter 4). The size of the proposed approach for developing the correlation network can rapidly grow as the size of the simulation model increases. Thus, the process of finding the common components of the event times for performing the subtraction and calculating the queue performance measures will become very time consuming. Therefore, in this implementation, the user can provide a limit for the number of generations for which the correlation network will be tracked for the fuzzy event times. In our experiments, we set the limit for the number of generations equal to 200. This limit will make the simulation process faster, but may decrease the accuracy of the results. In the future, the implications of limiting the number of generations on accuracy of the model should be further investigated.

The performance of the proposed FDES methodology has been tested using previously solved examples of fuzzy queueing systems found in the literature. Examples of single server queueing systems with fuzzy inter-arrival times and service times are used (Kao et al. 1999, Negi and Lee 1992). The estimated fuzzy numbers for average queue length and waiting time using the proposed methodology in this research adhere to the mathematically calculated results in all experiments. For example, a single server queueing system was analytically solved in (Kao et al. 1999), where the fuzzy inter-arrival time is A=trap(5,6,7,9) and the fuzzy service time is defined as S=trap(4,4.5,5.25,7.5). This example is modelled using the developed FDES template in Simphony.NET as illustrated in Figure 5.3.

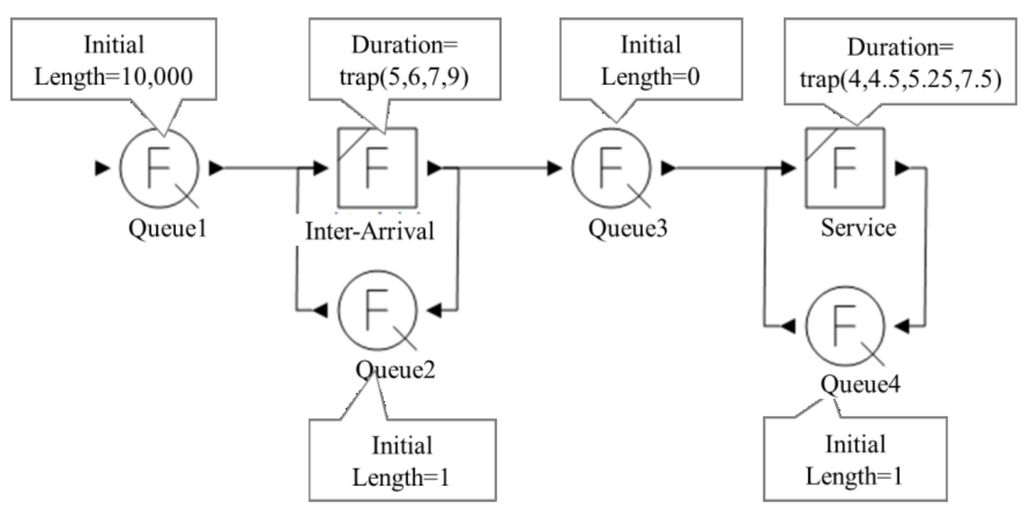


Figure 5.3 A queueing system modelled in FDES with fuzzy service time and fuzzy inter-arrival time

The results for L obtained from the FDES for 11 alpha-cuts are indicated in Table 5.1. These results are based on 10,000 entities (n=10,000). For the upper limits of the alpha-cuts, the results reported by Kao et al. (1999) are ∞. This means that a steady state cannot be reached for some of the alpha-cuts of the fuzzy queueing system. Inability to reach a steady state means that the queue size will increase as new entities arrive because entities will leave the queue with a smaller rate.

In the simulation model, the calculated values for the upper limits of L_α for $\alpha \leq 0.7$ will always increase by increasing the number of entities (e.g., n=20,000). This behaviour complies with the analytically calculated results (∞). Furthermore, the calculated values for the lower limits of L_α for all values of α and the upper limits of L_α for $\alpha > 0.7$ are equal to the analytically calculated results reported by Kao et al. (1999), as illustrated in Table 5.1.

Table 5.1 Alpha-cuts of L for the single server example with fuzzy service time and time between arrivals

Alpha level	*Alpha cuts for L*		*Alpha cuts for L*	
	FDES		*Kao et al. (1999)*	
	Lower	*Upper*	*Lower*	*Upper*
0.0	0.44	2501.24*	0.44	∞
0.1	0.46	2133.56*	0.46	∞
0.2	0.48	1780.02*	0.48	∞
0.3	0.49	1439.82*	0.49	∞
0.4	0.51	1112.22*	0.51	∞
0.5	0.53	796.53*	0.53	∞
0.6	0.55	492.12*	0.55	∞
0.7	0.57	198.38*	0.57	∞
0.8	0.59	0.98	0.59	0.98
0.9	0.62	0.92	0.62	0.92
1.0	0.64	0.87	0.64	0.87

(*) Value increases by increasing the number of entities in the simulation

The performance measures of fuzzy queueing systems that are calculated with FDES indicate complete compliance with the analytically solved performance measures. Although, analytical methods for solving fuzzy queueing systems are capable of solving simple fuzzy queueing systems, these approaches are tedious and restrictive in the type of queueing systems that they can solve (Munoz and Ruspini 2013, Negi and Lee 1992). On the other hand, the proposed extension to FDES allows calculating queue performance measures in practical simulation models of construction projects as illustrated in the next section.

5.5 PRACTICAL EXAMPLE OF ASPHALT PAVING OPERATION

A simplified example of an asphalt paving operation is provided to illustrate the practical aspects of the proposed approach for calculating average waiting time in FDES for construction management. For illustrative purposes, the duration of a paving activity is represented as a fuzzy number that is estimated by explicitly modelling two qualitative factors. Then the proposed methodology in FDES is used to analyze the queues in the asphalt paving operation and to find the optimum number of trucks.

5.5.1 Asphalt Paving Operation

In an asphalt plant, the trucks are loaded with the hot-mix asphalt. Each truck then hauls the mix material to the site. On the site, the haul truck waits for the first available paver. Then the truck is driven to the start location of the paving operation. The back of the haul truck is then aligned with the front of the paver. The truck dumps the mixture to the paver and returns to haul more asphalt mixture until all of the mixture is hauled to the site.

The pavers distribute the mixture through its screed. In order to make sure that the asphalt mixture is continuously placed, it is critical that the pavers do not remain idle. The compaction starts after spreading the mixture. The rollers move slowly on asphalt to achieve the desired level of compaction. It is critical for the roller to compact the mixture while its temperature is still above 85°C. As a result, the trucks cannot wait very long for the pavers, in order to prevent the mixture from cooling down before compaction. Quality control and quality assurance tests are often conducted after the compaction process (Hassan and Gruber 2008, Lu 2003, Nassar et al. 2003).

Many factors impact the duration of activities in the asphalt paving operation (Chio and Minchin 2006, Mostafavi et al. 2012). Fuzzy set theory and expert judgment can be used to estimate the duration of activities in the paving operation based on explicitly modelled factors. The next section provides an example of estimating the duration of the paving activity by explicitly considering two qualitative factors using fuzzy set theory.

5.5.2 Estimating Fuzzy Numbers for the Duration of Paving Activity

Generally, fuzzy numbers can be defined for activity durations in construction projects when expert judgment is used to estimate the durations. As discussed in Section 2, different approaches exist for developing fuzzy numbers for activity durations: (1) the fuzzy numbers can be directly derived from experts, or (2) the impact of factors that affect the activity durations can be expressed explicitly by experts. In this example, the duration of the paving activity is estimated as a fuzzy number by explicitly modelling a sample of factors that may impact the paving activity.

For illustrative purpose of this example, the impact of two factors, weather conditions and skill level of pavers' operators, are considered on the duration of the paving activity. Weather conditions and skill level are assessed qualitatively. Weather is classified into 3 linguistic terms: *good*, *poor*, and *average*. Skill level of pavers' operators is also classified into 3 linguistic terms: *low*, *high* and *medium.* The duration of the paving activity is classified into 5 linguistic terms: *very small*, *small*, *medium*, *large*, and *very large*.

If-then rules can be developed to relate the input factors to the duration of the activity (AbouRizk and Sawhney 1993, El-Rayes and Moselhi 2001, Pan 2001, Shaheen et al. 2009). Generally, the if-then rules should be developed (1) based on the extensive knowledge of an expert or a group of experts, or (2) by using historical data. For example, an expert may provide the following rule:

> If the weather is *good* and the operators' skill level is *high*, then the duration of the paving activity is *very small*.

A set of if-then rules that relate weather conditions and skill level of operators to the duration of the paving activity is defined in Table 5.2. The developed if-then rules in this table are for illustrative purposes only.

Table 5.2 If-then rules for the duration of paving activity

Weather conditions	*Skill level of pavers' operators*	*Duration of paving activity*
Good	*High*	*Very small*
Good	*Medium*	*Small*
Good	*Low*	*Medium*
Average	*High*	*Medium*
Average	*Medium*	*Medium*
Average	*Low*	*Large*
Poor	*High*	*Large*
Poor	*Medium*	*Very large*
Poor	*Low*	*Very large*

A fuzzy number is defined for each of the linguistic terms of the duration of paving activity to provide a mathematical representation for that term. The fuzzy number of each linguistic term is developed based on the minimum (*min*) and maximum (*max*) possible value of the duration (Ramze et al. 1999). First, Equation 5.14 is used to calculate m, m_1, and m_2.

$$m=(min+max)/2$$

$$m_1=(3min+max)/4$$

$$m_2=(min+3max)/4 \quad (5.14)$$

Then the triangular fuzzy numbers are developed for each of the linguistic terms of paving activity according to Equation 5.15. These fuzzy numbers are also illustrated in Figure 5.4.

$$\textit{Very small} \text{ duration} = \text{tri}(min,min,m1)$$

$$\textit{Small} \text{ duration} = \text{tri}(min,m1,m)$$

$$\textit{Medium} \text{ duration} = \text{tri}(m1,m,m2)$$

$$\textit{Large} \text{ duration} = \text{tri}(m,m2,max)$$

$$\textit{Very large} \text{ duration} = \text{tri}(m2,max,max) \quad (5.15)$$

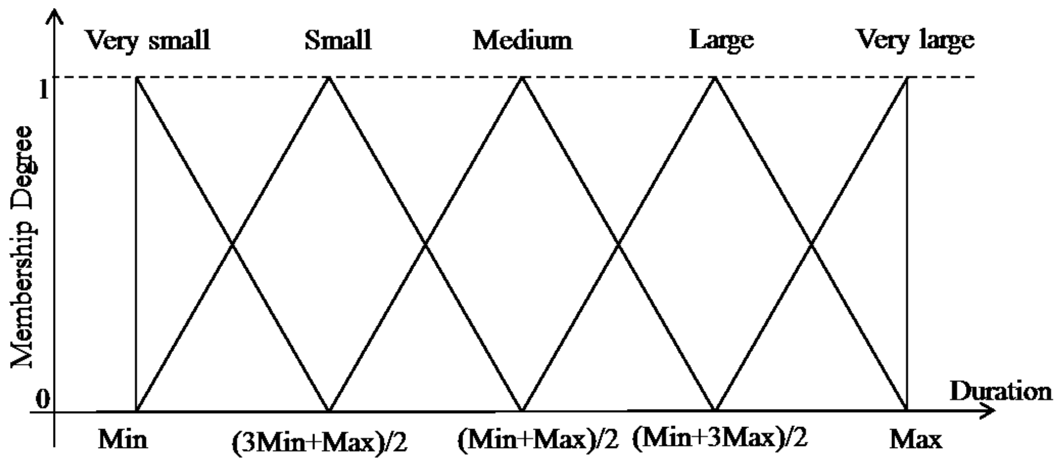


Figure 5.4 An example of defining fuzzy numbers for an activity with 5 linguistically expressed categories

The minimum and maximum value of the duration of the paving activity for paving one segment of road can be estimated based on the length of the segment and minimum and maximum possible speed of paver. The minimum and maximum possible paving speed can be defined based on expert(s) judgment. In this example, the length of one segment is assumed to be 120 meters and the minimum and maximum paving speed is assumed to be 20 metres/minute and 40 metres/minute. Thus, the minimum and maximum values for the activity durations of the paving activity for one segment of asphalt are estimated as 3 and 6 minutes. Therefore, according to equations 5.14 and 5.15, the fuzzy numbers for *Very small* duration, *Small* duration, *Medium* duration, *Large* duration, and *Very large* duration are tri(3,3,3.75), tri(3,3.75,4.5), tri(3.75,4.5,5.25), tri(4.5,5.25,6), and tri(5.26,6,6), respectively. The above approach provides an intuitive way of defining the fuzzy numbers for the activity duration. However, the assumed fuzzy numbers are for illustrative purposes only, and the fuzzy numbers can be best defined for a specific project based on expert judgment.

The developed if-then rules can be used to estimate the fuzzy number for the duration of the paving activity based on different combinations of weather conditions and operators' skill level. The fuzzy number for the duration of the paving activity is provided as input to a FDES model of the asphalt paving operation. The proposed methodology of calculating average waiting time is then used to find the optimum number of trucks in the asphalt paving operation as discussed in the next section.

5.5.3 Finding Optimum Number of Trucks in Asphalt Paving Operation using FDES

Simulation of an asphalt paving operation helps to calculate the number of required resources to maintain the essential balance between the production and placement of hot mix asphalt (Starry 2009). Various discrete event simulation models have been developed for simulating an asphalt paving operation (Hassan and Gruber 2008, Lu 2003, Nassar et al. 2003, Zhang et al. 2014, Labban et al. 2013). However, all previous simulation models use traditional DES and only assume stochastic uncertainty (represented by probability distribution) for activity durations. On the other hand, using a FDES framework allows the consideration of subjective uncertainty (represented by fuzzy numbers) in the activity durations of the asphalt paving operation.

A simulation model is developed to estimate the average queue length and waiting time of trucks for pavers and the idle time of pavers (waiting time of pavers for trucks) in the asphalt paving process, using the FDES framework. For simplicity of this example, only the asphalt paving loop is analyzed. In the developed model, one surge bin in the asphalt plant loads the trucks with the hot-mix asphalt. The trucks then haul the asphalt mix to the construction site. On the site, the back of the truck is aligned to the first available paver to dump the hot-mix asphalt to the paver. The paver then paves one segment of the road, which is assumed to be 120 meters. Two pavers are assumed in this example. The developed FDES template in Simphony .NET is used to develop this simulation model (Figure 5.5).

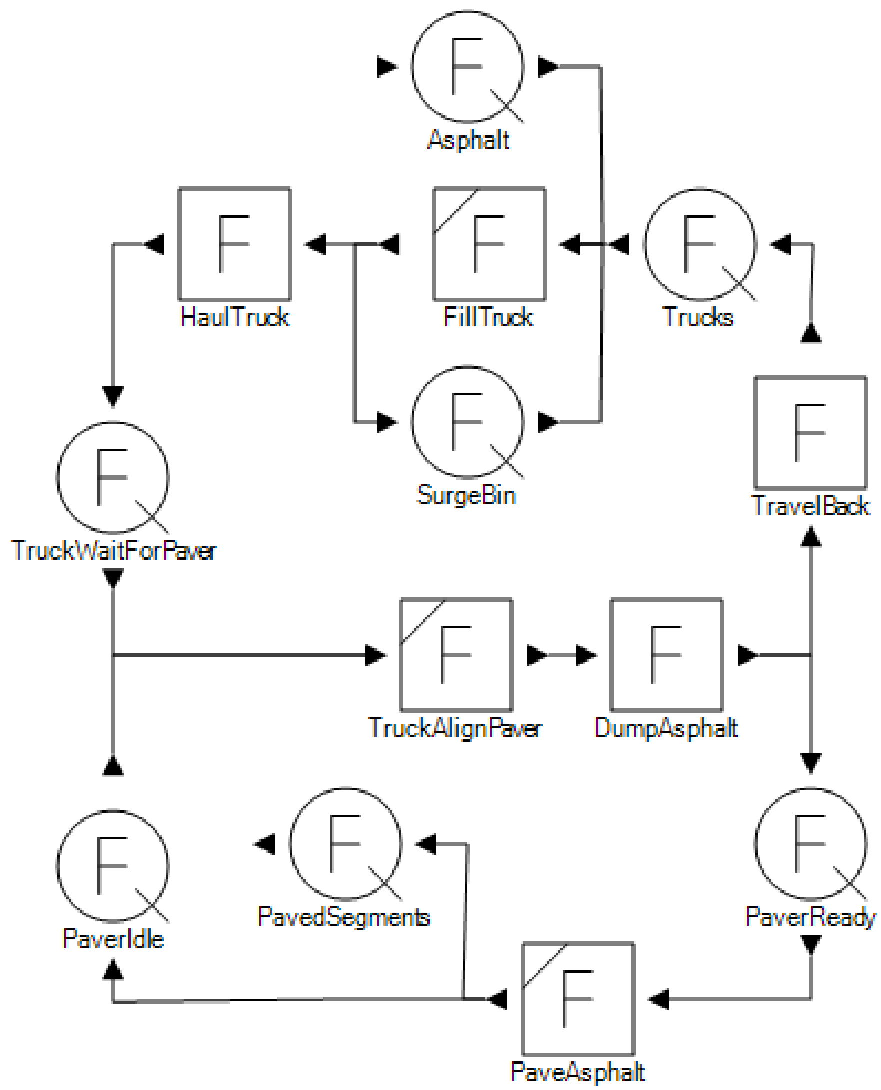


Figure 5.5 Simulation model developed for paving operation using FDES template

The duration of the paving activity is assumed to be a fuzzy number, as discussed in the previous section, derived based on if-then rules that relate different combinations of weather conditions and operators' skill level to duration. The duration of other activities are adapted from Brock (2014) (Table 5.3).

Table 5.3 Activity durations of asphalt paving operation

Activity	*Duration (minutes)*
Load truck at plant	1
Truck travel to site	20
Truck aligned with paver	2
Truck dump asphalt to paver	3
Truck travel back to plant	20

Considering the relatively high cost of pavers compared to trucks and the necessity of having a continuous paving operation to achieve the optimum results, the number of trucks associated with zero average idle time for pavers (waiting time of pavers for trucks) should be found. The average waiting time of trucks for pavers should be minimized as the second objective of this problem (Hassan and Gruber 2008). The proposed methodology (Section 4) will allow us to find the average idle time of pavers (waiting time of pavers for trucks) and waiting time of trucks for pavers in the FDES developed for the asphalt paving process (Figure 5.5). The simulation model should be run for different numbers of trucks, and the number of trucks associated with zero average idle time of pavers and minimum waiting time for trucks should be chosen as the optimum number.

The optimum number of trucks is sensitive to the qualitative factors (weather conditions and operators' skill level) that are used to estimate the duration of the paving activity. For example, if the weather conditions is *poor* and the operators' skill level is *high*, the duration of the paving activity is *large*, according to Table 5.2. Thus, the duration of paving is tri(4.5, 5.25, 6). The FDES model (Figure 5.5) is used to model the paving operation for different numbers of trucks. Table 5.4 indicates the average waiting time of trucks and idle time of pavers for this scenario calculated using the FDES model. The calculated values are triangular fuzzy numbers. However, when the parameters a, b, and c of a triangular fuzzy number, tri(a,b,c), are equal, the fuzzy number can be represented as a crisp value. This is why some of values in Table 5.4 are crisp. According to Table 5.4, the number of trucks should be higher than 10 to have an average pavers' idle time equal to zero. Since the average waiting time of trucks has to be minimized as the second objective, 10 trucks will be the optimum number of trucks for this example. In this case, the average waiting time (W_q) of trucks is a triangular fuzzy number tri(1.53, 5.02, 8.56). Also, the average queue length (L_q) of the trucks on the site is the fuzzy number represented in Figure 5.6. This fuzzy number does not have a standard shape such as triangular or trapezoidal. The values in this fuzzy number range from 0.6 to 3.6.

On the other hand, if weather conditions is *good* and the skill level is *high*, the duration of the paving activity is *small*; tri(3,3,3.75). For different numbers of trucks in this scenario, the developed model based on FDES is again used to calculate the average waiting time of trucks and idle time of pavers. In this case, according to the results of the simulation as illustrated in Table 5.5, the average idle time of pavers is equal to 0 when at least 12 trucks are available. Considering the second objective of minimizing the waiting time of trucks, 12 trucks is the optimum number of truck in this situation, and the average waiting time (W_q) of trucks is tri(1.5, 1.5, 6.02). Also, the average queue length (L_q) of trucks on the site is a triangular fuzzy number tri(0.32, 0.32, 1.30).

Table 5.4 Average waiting time of trucks and idle time of pavers for different number of trucks when weather conditions is poor and skill level of pavers operators is high

Number of trucks	*Average idle time of pavers*	*Average waiting time of trucks*
8	tri(0.50,1.24,1.9)	tri(0.01,0.01,0.02)
9	tri(0,0,0.72)	tri(0.03,0.16,3.51)
10	0.00	tri(1.53,5.02,8.56)
11	0.00	tri(6.35,9.88,13.9)
12	0.00	tri(10.50,14.99,19.04)
13	0.00	tri(15.23,20.09,24.95)

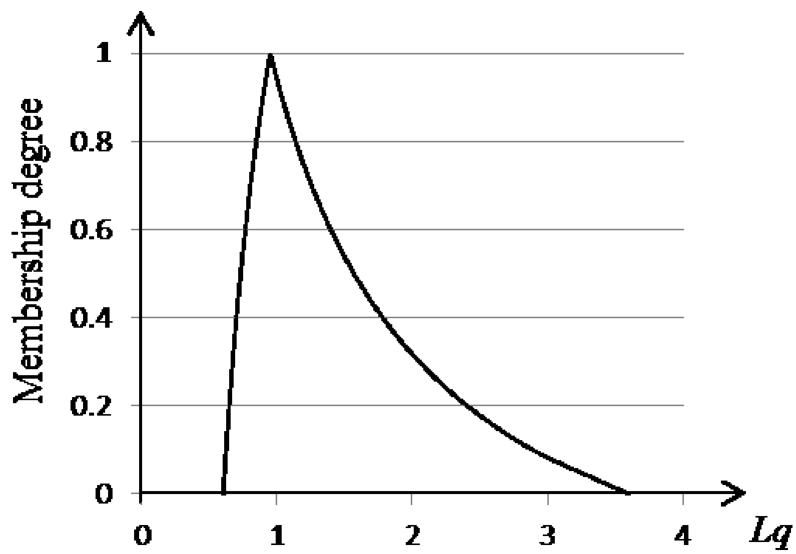


Figure 5.6 Average queue length (*Lq*) of trucks waiting for pavers when there are 10 trucks, weather conditions is *poor*, and skill level of pavers' operators is *high*

Table 5.5 Average waiting time of trucks and idle time of pavers for different number of trucks when weather conditions is average and the skill level of pavers operators is medium.

Number of trucks	*Average idle time of pavers*	*Average waiting time of trucks*
8	tri(2.47,3.49, 3.49)	0.01
9	tri(1.47,2.23, 2.23)	0.02
10	tri(0.45,1.19, 1.19)	0.03
11	tri(0.00,0.36, 0.36)	tri(0.04,0.04, .16)
12	0.00	tri(1.5, 1.5, 6.02)
13	0.00	tri(5.53,5.53,0.32)

In summary, in this example:

1. Fuzzy set theory is used to mathematically model the impact of qualitative factors on the duration of the paving activity using if-then rules. Thus, a fuzzy number is estimated for the duration of the paving activity.
2. FDES is used to simulate the asphalt paving process in which the duration of the paving activity is expressed as a fuzzy number.
3. The proposed methodology for FDES is used to calculate average queue length and waiting time of trucks on the site, as well as idle time of pavers; allowing us to find the optimum number of trucks for the asphalt paving activity.

The optimum number of trucks is sensitive to the linguistic values of the qualitative factors (weather conditions and skill level of pavers' operators). Thus, the proposed approach enables the explicit consideration of these qualitative factors on the optimum number of trucks for the asphalt paving operation.

5.6 CONCLUDING REMARKS

Previously developed FDES frameworks only calculate simulation time (e.g. project completion time) for the simulation output; they do not have the capability of calculating queue performance measures such as average queue length and waiting time, both of which are important for decision making in construction projects. This chapter proposes an approach for calculating average queue length and waiting time in FDES. First, a correlation network is developed to track the correlation of event times. The subtraction of fuzzy event times for calculation of average queue length and waiting time is then performed by considering the developed correlation network. The proposed approach is validated using analytically solved queuing examples. The practicality of the proposed approach is illustrated using an asphalt paving operation in which the number of trucks has been optimized. The methodology presented in this chapter for calculating average queue length and waiting time in FDES will broaden the practical benefits of FDES in construction project analysis, allowing us to perform what-if scenarios and optimization of construction resources.

In DES, only probability distributions can be used to represent the uncertainty of activity durations; On the other hand, in FDES, fuzzy numbers can be used to represent the uncertainty of activity durations. However, fuzzy numbers are appropriate for representing subjective uncertainty, while probability distributions are able to model stochastic uncertainty. Thus, subjective and stochastic uncertainty may simultaneously exist in a model (Zadeh, 2008). For example, in the asphalt paving simulation example, enough historical data may be available for the traveling time of trucks, and thus truck traveling times may be estimated using a probability distribution. On the other hand, the duration of paving activity may be estimated using a fuzzy number due to the unavailability of data and the qualitative assessment of the impact of factors on the duration of paving activity. Therefore, both fuzzy numbers and probability distributions simultaneously exist in the simulation, requiring a simulation framework that can handle both types of uncertainty. In the next chapter of this dissertation, a hybrid simulation framework will be developed for construction management that can simultaneously handle both fuzzy numbers and probability distributions.

5.7 REFERENCES

AbouRizk, S. (2010). Role of simulation in construction engineering and management. Journal of Construction Engineering and Management, 136(10), 1140-1153.

AbouRizk, S. M., & Sawhney, A. (1993). Subjective and interactive duration estimation. Canadian journal of civil engineering, 20(3), 457-470.

Adeli, H., & Hung, S. L. (1994). Machine learning: neural networks, genetic algorithms, and fuzzy systems. John Wiley & Sons, Inc.

Adeli, H., & Sarma, K. C. (2006). Cost optimization of structures: fuzzy logic, genetic algorithms, and parallel computing. John Wiley & Sons.

Brock , D. (2014). Hot mix asphalt trucking, Technical Paper, www.roadtec.com, Web. 23 Oct. 2014.

Carlsson, C., & Fullér, R. (2004), Additions of Completely Correlated Fuzzy Numbers, IEEE International Conference on Fuzzy Systems, Budapest, Hungary, pp. 535-539.

El-Rayes, K., & Moselhi, O. (2001). Impact of rainfall on the productivity of highway construction. Journal of construction engineering and management, 127(2), 125-131.

Halpin, D. W. (1992). Planning and analysis of construction operations. John Wiley & Sons.

Hassan, M. M., & Gruber, S. (2008). Application of discrete-event simulation to study the paving operation of asphalt concrete. Construction Innovation: Information, Process, Management, 8(1), 7-22.

Kao, C., Li, C. C., & Chen, S. P. (1999). Parametric programming to the analysis of fuzzy queues. Fuzzy Sets and Systems, 107(1), 93-100.

Labban, R., AbouRizk, S., Haddad, Z., & Elsersy, A. (2013). A discrete event simulation model of asphalt paving operations. Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World, Washington, DC, pp. 3215-3224, IEEE Press.

Little, J. D. C. (1961). "A Proof for the Queuing Formula: L = λW". Operations Research 9 (3): 383–387.

Lu, M. (2003). Simplified discrete-event simulation approach for construction simulation. Journal of Construction Engineering and Management, 129(5), 537-546.

Mostafavi, A., Valentin, V., Abraham, D. M., & Louis, J. (2012). Assessment of the productivity of nighttime asphalt paving operations. Journal of Construction Engineering and Management, 138(12), 1421-1432.

Nassar, K., Thabet, W., & Beliveau, Y. (2003). Simulation of asphalt paving operations under lane closure conditions. Automation in construction, 12(5), 527-541.

Negi, D. S., & Lee, E. S. (1992). Analysis and simulation of fuzzy queues. Fuzzy sets and systems, 46(3), 321-330.

Pan, N. F. (2005). Assessment of productivity and duration of highway construction activities subject to impact of rain. Expert systems with Applications, 28(2), 313-326.

Perrone, G., Zinno, A., & La Diega, N. (2001). Fuzzy discrete event simulation: A new tool for rapid analysis of production systems under vague information. Journal of Intelligent Manufacturing, 12(3), 309-326.

Ramze Rezaee, M., Goedhart, B., Lelieveldt, B. P., & Reiber, J. H. (1999). Fuzzy feature selection. Pattern Recognition, 32(12), 2011-2019.

Shaheen, A. A., Fayek, A. R., & AbouRizk, S. M. (2009). Methodology for integrating fuzzy expert systems and discrete event simulation in construction engineering. Canadian Journal of Civil Engineering, 36(9), 1478-1490.

Song, L., & Eldin, N. N. (2012). Adaptive real-time tracking and simulation of heavy construction operations for look-ahead scheduling. Automation in Construction, 27, 32-39.

Song, L., Ramos, F., & Arnold, K. (2008). A framework for real-time simulation of heavy construction operations. Proceedings of the 40th Conference on Winter Simulation, Austin, TX , pp. 2387-2395, IEEE.

Zadeh, L. A. (2008). Is there a need for fuzzy logic? Information Sciences,178(13), 2751-2779.

Zeng, Z., Xu, J., Wu, S., & Shen, M. (2014). Antithetic Method-Based Particle Swarm Optimization for a Queuing Network Problem with Fuzzy Data in Concrete

Transportation Systems. Computer-Aided Civil and Infrastructure Engineering, 29(10), 771-800.

Zhang, M., Heitzman, M., & Smith, A. E. (2013). Improving Hot Mix Asphalt Production Using Computer Simulation and Real Time Optimization. Journal of Computing in Civil Engineering, 28(3).

CHAPTER 6 HYBRID FUZZY DISCRETE EVENT SIMULATION (HFDES) FRAMEWORK

6.1 INTRODUCTION

The input parameters of construction simulation models are often uncertain. This uncertainty is modelled with probability distributions in traditional discrete event simulation (DES). On the other hand, as discussed in previous chapters, fuzzy discrete event simulation (FDES) provides a methodology for handling fuzzy uncertainty of the activity durations in DES.

Activity Durations are one of the most difficult parameters of the simulation models to estimate. Activity durations can be the direct input to the simulation model (Figure 6.1(a)), or they can be calculated using another component that is integrated with the simulation model. For example, an arithmetic component can be used to estimate the activity duration based on the productivity, number of workers, quantity, and shift duration (Figure 6.1(b)). Due to the high impact of different influencing factors (e.g. weather and skill level) on productivities of construction activities, a productivity prediction model may be also integrated with the simulation model to predict the duration by explicitly modelling various influencing factors (Figure 6.1(c)). In this scenario, the modelled influencing factors will be the indirect input to the simulation model. The uncertainty of some of these factors can be defined as probability distributions; on the other hand, other factors can be defined with fuzzy numbers to represent subjective uncertainty.

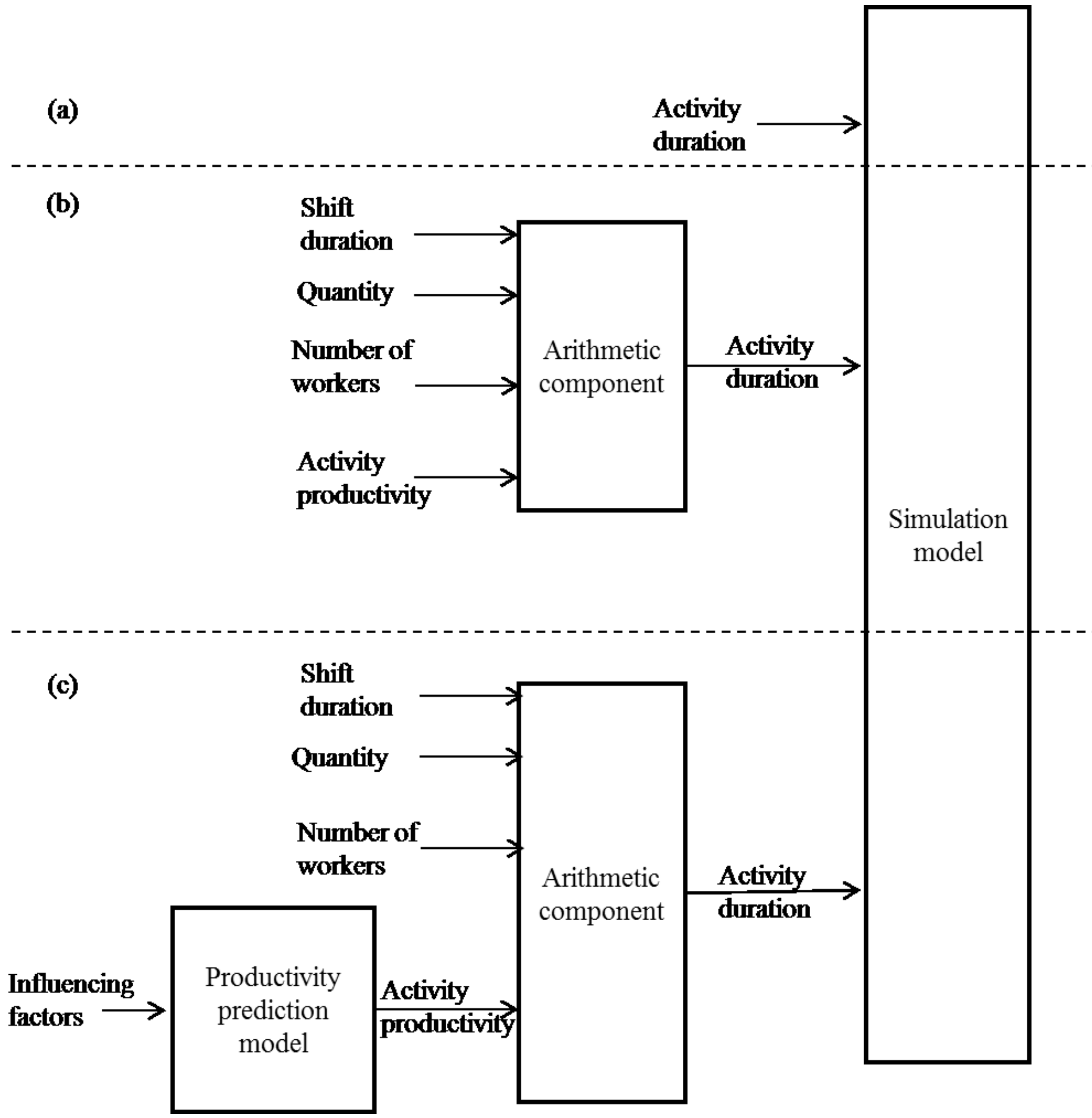


Figure 6.1 Activity duration as inputs to a construction simulation model (a) the activity duration is a direct input to the model (b) the activity duration is estimated from productivity, shift duration, quantity, and number of workers (c) the productivity is estimated using a productivity prediction model and then used to estimate the activity duration

The decision on the method for representing the uncertainty of direct or indirect input parameters of a simulation model depends on the availability of data and nature of each input variable. Thus, both fuzzy numbers and probability distributions may be used for representing the uncertainty of the input parameters, requiring a hybrid simulation framework that can handle both types of uncertainties simultaneously.

In recent years, many researchers in various areas proposed hybrid approaches for considering both fuzzy and stochastic uncertainty (e.g. Baudrit et al. 2008, Chen et al. 2003, Chen et al. 2010, , Cooper et al. 1996, Davis and Keller 1997, Huang 1998, Liu et al. 2003, Möller and Beer 2004, Sadeghi et al. 2010). However, no framework is yet proposed that can analyze both fuzziness and randomness (i.e. stochastic uncertainty) in DES.

In this chapter, a hybrid discrete event simulation (HFDES) framework that can simultaneously handle fuzzy and stochastic uncertainty is proposed. Furthermore, a methodology for interpretation of the results is proposed for construction management. This chapter is organized as follows: Section 6.2 provides an approach to help an investigator choose an appropriate method for modelling the simulation inputs. Section 6.3 proposes the HFDES approach for dealing with simultaneous fuzzy and stochastic uncertainty. Sections 6.4 and 6.5 provide methodologies for analyzing the output of HFDES. Section 6.5, compares the results of HFDES framework with analytically solved queueing examples. Section 6.7 illustrates the practicality of HFDES using a real case study of module assembly process.

6.2 REPRESENTING THE UNCERTAINTY OF SIMULATION INPUTS

Figure 6.2 illustrates the flowchart that relates the information availability and nature of a parameter to the method of representing the uncertainty (Dubois and Guyonnet 2011).

In this flowchart, a series of questions are asked from the investigator to guide her/him to the final representation of an input parameter. First, the investigator is asked whether the input parameter should be represented as a variable or a deterministic value that is not subject to variability. If a parameter is considered to have a deterministic nature (e.g. depth of a well, length of a spool), questions are asked regarding the information of the investigator regarding that parameter:

1) If the value is precisely known, the parameter is presented as a crisp value.
2) If the investigator can only provide an interval for the variable, a simple interval [a,b] is used to represent the variable. An interval can be assumed a special case of a fuzzy number.
3) If the investigator can express preference in the interval, a fuzzy set is developed for the variable.

On the other hand, if the parameter is considered to have a variable nature, the investigator is asked the following questions:

1) If statistical data are available for a variable, a probability distribution is developed for that variable.
2) If statistical data are not available, but the type of distribution is known, a probability distribution can be elicited from expert judgment. However, this probability elicited from expert judgment usually contains some imprecision in its parameters; imprecise probabilities can be used to represent the uncertainty in these situations. Using imprecise probability distributions as the inputs of the

simulation model is outside of the scope of this thesis and is recommended for future research.

3) If no data are available and the type of the distribution of the parameter is not known, a fuzzy set or interval is developed for the parameter based on expert judgment.

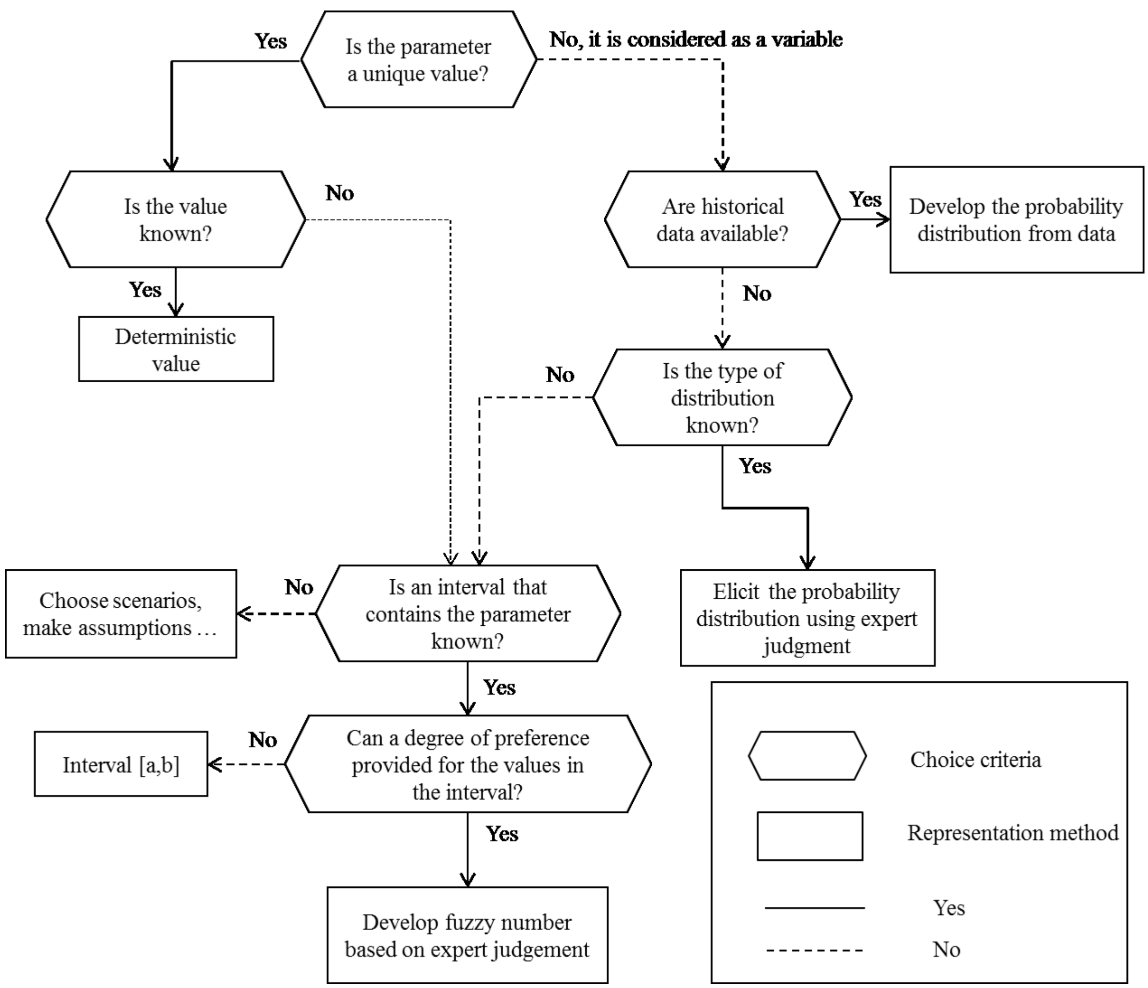


Figure 6.2 Flowchart relating the nature of information to the method of representation of uncertainty (adapted from Dubois and Guyonnet 2011)

Using the above methodology, some input parameters to a simulation model may be represented with probability distributions and some parameters may be represented with fuzzy numbers. Furthermore, when a prediction model is used to predict the productivity of construction projects (Figure 6.1(c)), the inaccuracy of the predicted productivity may be represented using a fuzzy number or a probability distribution depending on the nature of the model and availability of test data (As discussed in Chapters 2 and 3). In the next section, the methodology for processing both fuzzy and probabilistic uncertainty in a simulation model is proposed.

HYBRID FUZZYDISCRETE EVENT SIMULATION (HFDES) FRAMEWORK

The HFDES framework can process both fuzzy numbers and probability distributions (i.e. random variables) as the inputs of the simulation model. Consider probability distributions, $R_1, R_2 \ldots R_n$, and fuzzy numbers, $F_1, F_2 \ldots F_m$, as the direct inputs to a simulation model, $M(R_1, R_2 \ldots R_n, F_1, F_2 \ldots F_m)$.

In the proposed HFDES framework, sample sets are produced from the probability distributions. For each sample set #i, a crisp value is assigned to each of the input parameters that are modelled with probability distributions, $r_{1i}, r_{2i} \ldots r_{ni}$. Therefore, in each sampling, the model can be presented as $M(r_{1i}, r_{2i} \ldots r_{ni}, F_1, F_2 \ldots F_m)$. Genrally crisp values can be assumed as special case of fuzzy number; therefore, model M contains only fuzzy input variables.

If some of the parameters $r_{1i}, r_{2i} \ldots r_{ni}, F_1, F_2 \ldots F_m$ are indirect parameters that are used for estimating the durations of activities in the simulation model (e.g. Figure 6.1(b and c)), fuzzy arithmetic should be first used to calculate the fuzzy activity duration, this is because, prediction models such artificial neural networks or fuzzy rule-based systems are based on arithmetic operations. As discussed in Chapter 4, fuzzy arithmetic allows calculating the output of any function as a fuzzy number when the inputs are fuzzy numbers. After calculating the activity durations as a fuzzy number, FDES is employed to process fuzzy numbers in the simulation as discussed in Chapters 4 and 5. This procedure is illustrated in Figure 6.3. The number of samplings from the random variables is presented as *w* in this figure.

Different types of outputs can be obtained from DES of construction projects such as the completion time of the project, man-hours spent on the project, queue waiting times, etc. In a FDES, these outputs are fuzzy numbers as discussed in Chapters 4 and 5. For each type of output, *w* fuzzy numbers are obtained in the HFDES framework as indicated in Figure 6.1. In the next section, a method for analyzing and making decisions with these outputs is provided.

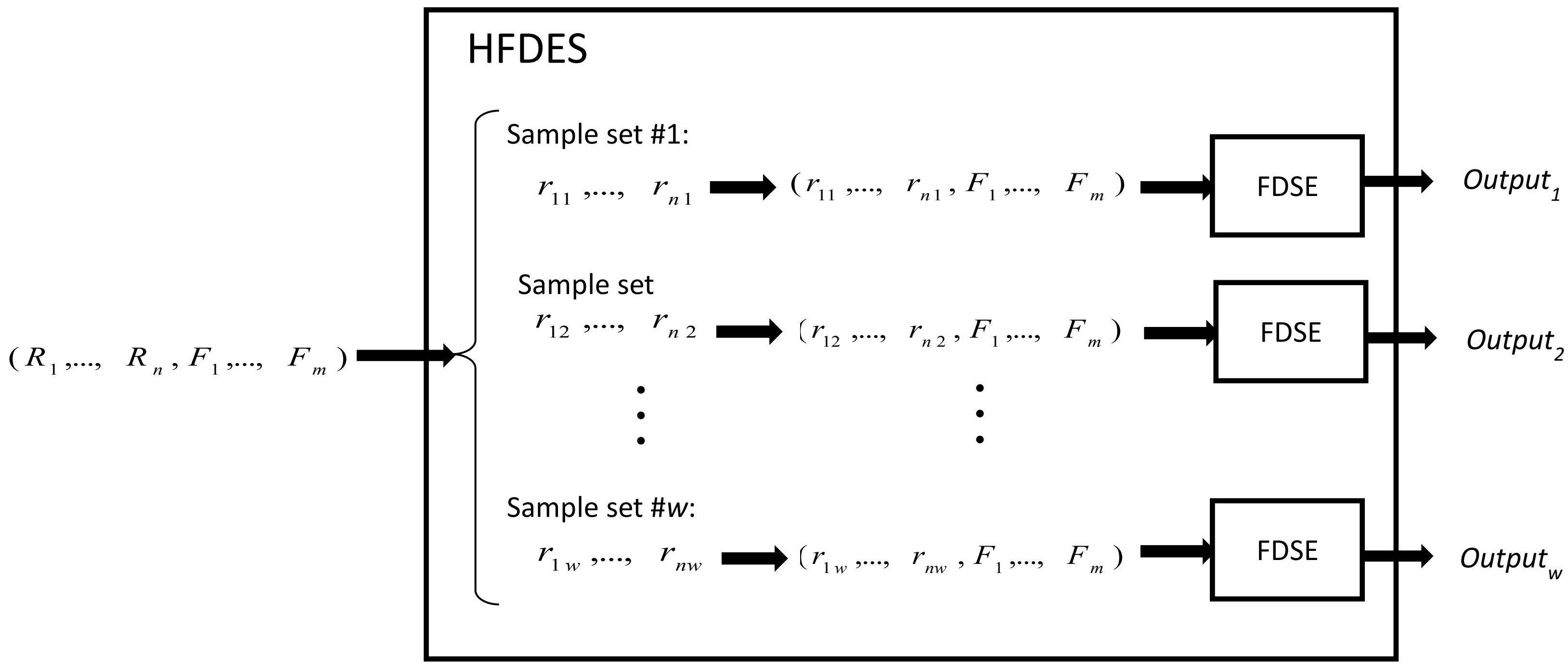


Figure 6.3 Hybrid Discrete Event Simulation (HFDES) Framework

6.3 INTERPRETING THE OUTPUT OF HFDES

In traditional DES, the output is usually represented as a distribution for decision-making purposes. Each of the outputs in traditional DES is a crisp value; and a PDF or a cumulative distribution function (CDF) is fitted over these outcomes. CDF is typically used in decision-making and reliability analysis in construction management. It allows one to find for the probability of not exceeding a given threshold. For example, in construction management, decision makers are often interested in finding the probability that a project will be completed within a certain value of cost or time. Equation 6.1 defines the CDF function of a random variable X (Ahuja and Nandakumar 1985).

$$F_X(x) = \Pr\{X \leq x\} \tag{6.1}$$

The inverse of the CDF is the quantile function. The quantile function, $F_X^{-1}(p)$, is used to calculate the value, x, so that the output will be equal or less than x with the provided probability, p. Quantile function is commonly used for bidding and decision making in construction management where p represents the confidence level for that decision. For example, one may want to estimate the completion time of a project with 95% confidence. This value is referred as the 95th quantile of the output. In the context of the simulation process, 95% of the conducted simulation results are less than the 95th quantile of the output. Considering a finite number of outputs of a traditional DES, the CDF function of the output can be estimated using Equation 6.2, where w is the number of outputs of the simulation model.

$$F_{Output}(x) = \frac{\text{Number of outputs that are less than or equal to } x}{w} \tag{6.2}$$

In the case of HFDES, each of the output samples of the simulation is a fuzzy number. Having a number of measurements or simulation outputs, one can develop a fuzzy CDF from those outputs (Sadeghi et al. 2010, Kentel and Aral 2005). The fuzzy CDF is a CDF that contains fuzziness. Alpha-cut approach can be used to develop a fuzzy CDF: first the infimum and supremum values of the alpha-cuts of the samples are used to calculate two CDFs at each alpha level, $F_{\propto,inf}$ and $F_{\propto,sup}$. Second, $F_{\propto,inf}$ and $F_{\propto,sup}$ will generate a CDF bound $F_{\propto}(x)$ at each alpha level (Figure 6.4).

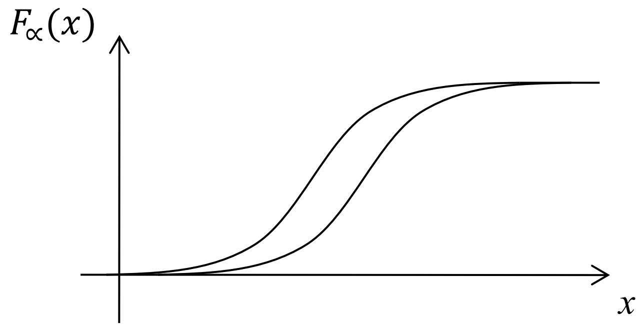


Figure 6.4 PDF bound represents a range of CDFs in an alpha level (Adapted from Sadeghi et al. 2010)

The final fuzzy CDF, $F(x)$, can be determined by aggregating CDF bounds at different levels of alpha based on the representation theorem (Equation 6.3; Pedrycz and Gomide, 2007).

$$F(x) = Sup_{\alpha\in[0.1]}[\alpha F_{\alpha}(x)] \tag{6.3}$$

Fuzzy CDF, $F(x)$, represents both probability and membership degree. Therefore, a three dimensional graph should be used to represent both probability and membership degree of each value. Containing both fuzzy and stochastic uncertainty in a fuzzy CDF makes its interpretation difficult. For example, in the fuzzy CDF, the probability of getting an output less than a specific threshold is a fuzzy number. Figure 6.5 indicates an example of a fuzzy CDF in which the probability of obtaining a value less than 1.1 is calculated (Sadeghi et al. 2010). The calculated confidence level in this fuzzy CDF is a fuzzy number that provides a membership degree for different probability values.

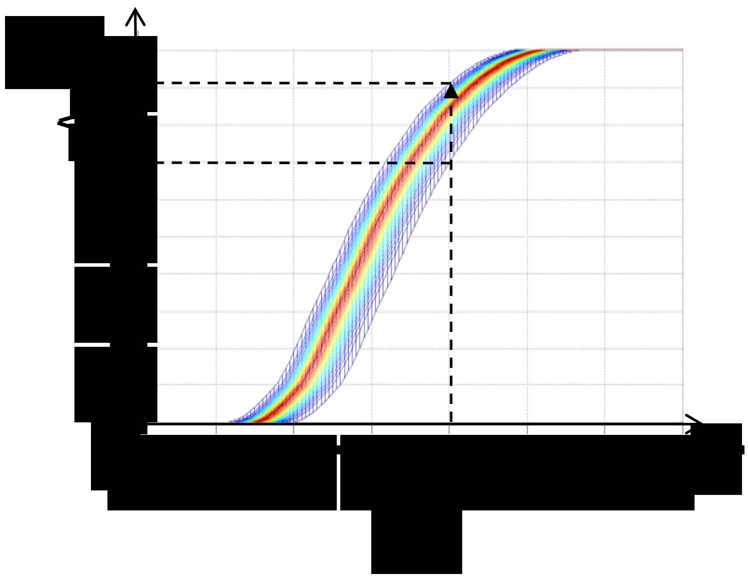


Figure 6.5 Calculating the probability that value x is smaller than 1.1 (a) fuzzy CDF (b) fuzzy number representing the probability of getting a value less than 1.1 (Adapted from Sadeghi et al. 2010)

However, in construction projects, managers are often interested in a final crisp value representing the confidence level for evaluating the situation and making decisions. One may suggest using the most pessimistic value of the range as the final value of the confidence level. For example in Figure 6.5, the final confidence level will be 0.7 which is the lowest value in the estimated fuzzy number in Figure 6.5(b). However, this approach provides a very conservative estimate of the confidence level.

Dubois and Guyonnet (2011) pointed out the problem of decision making based on the results of hybrid approaches. Dubois and Guyonnet (2011) proposed using the credibility of an event as the "confidence index" of that event. The credibility (Cr) of an event *A* can be calculated using the average of possibility, *Pos*, and necessity, *Nec,* of that event (Equation 6.4).

$$\mathrm{Cr}\{A\} = \frac{Pos\{A\}+Nec\{A\}}{2} \tag{6.4}$$

Let *X* be a fuzzy number with membership function μ, the possibility and necessity of $\{X<x\}$ can be calculated according to equations 6.5, 6.6, respectively (Pedrycz and Gomide 2007).

$$Pos\{X \leq x\} = sup_{X\leq x}\mu(x) \tag{6.5}$$

$$Nec\{X \leq x\} = 1 - Pos\{X > x\} \tag{6.6}$$

For example, in a triangular membership function, *tri(a,b,c)* with membership function μ, if x<b, the $Pos\{X \leq x\}$ is equal to μ (x) and the $Nec\{X \leq x\}$ is equal to 0 (Figure 6.6(a)). Thus the $Cr\{X \leq x\}$ can be calculated according to Equation 6.7.

$$Cr\{X \leq x\} = \frac{\mu(x)}{2} \tag{6.7}$$

On the other hand, if x>b, the $Pos\{X \leq x\}$ is always 1 and the $Nec\{X \leq x\}$ is $1 - \mu(x)$ (Figure 6.6(b)). Thus, the $Cr\{X \leq x\}$ can be calculated according to Equation 6.8.

$$Cr\{X \leq x\} = \frac{2-\mu(x)}{2} \tag{6.8}$$

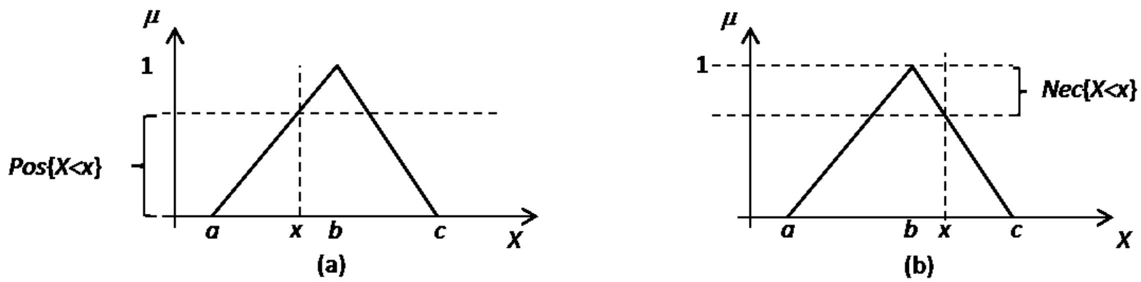


Figure 6.6 Illustrating possibility and necessity of X≤x (a) Illustrating Pos{X≤x}, Nec{X≤x}=0 (b) Illustrating Nec{X≤x}, Pos{X≤x}=1

Dubois and Guyonnet (2011) proposed developing a function $F_X(x)$ to represent the confidence index (CI) of $\{X \leq x\}$ when facing hybrid fuzzy and stochastic models. In a traditional DES, each simulation output (i.e. $Output_1$, $Output_2$...$Output_w$) is either smaller or larger than a threshold x. Thus, Equation 6.2 can be employed to calculate the probability of $\{Output \leq x\}$.

In HFDES, each output (i.e. $Output_1$, $Output_2$... $Output_w$), is a fuzzy number and a credibility degree can be assigned to whether an output is smaller or equal to a threshold x, $Cr\{Output_i \leq x\}$. Therefore, Equation 6.2 can be generalized to Equation 6.8 to calculate a function $F_{Output}(x)$ representing the confidence index of $Output \leq x$, $F_{Output}(x) = CI\{Output \leq x\}$ (Equation 6.9).

$$F_{Output}(x) = CI\{Output \leq x\} = \frac{\sum_{i=1}^{w} Cr\{Output_i \leq x\}}{w} \tag{6.9}$$

Using Equation 6.9, a single function, $F_{Output}(x)$, can be developed for representing the output of HFDES that can be employed for decision-making. $F_{Output}(x)$ is the same as a CDF developed for the output of a traditional DES, the only difference is that the y axis is defined as the "Confidence Index" in the hybrid fuzzy and random context, rather than the probability value in the probabilistic context.

Although, providing the range of the output of the simulation model as discussed above is important for decision-making purposes, in some situation the average value of the output of the simulation model may be required. For example, when validating the simulation model, the average value can be calculated and compared with the actual results. Also, calculating the fuzzy average waiting time or queue length in HFDES allows comparing the results with the analytically solved examples.

Having a number of fuzzy measurements, fuzzy arithmetic can be employed to calculate the mean value of the output, $\overline{output}$ of the simulation model from w samples according to Equation 6.10. Using this equation $\overline{output}$ is calculated as a fuzzy number using fuzzy addition, $\oplus$, and fuzzy division, $\oslash$ (Terán, 2007).

$$\overline{output} = (output_1 \oplus output_2 \oplus \ldots \oplus output_w) \oslash w \tag{6.10}$$

For example, in a queueing system, different performance measures such as W, W_q, L, L_q may be calculated using the FDES for each sample set. As discussed in Chapter 5, each of these measures are a fuzzy number in FDES. In order to find the performance measures of the hybrid system, fuzzy arithmetic can be employed. The results obtained from various samplings are averaged to find the final average performance measures of the queuing system, W, W_q, L, L_q (Equation 6.10). Thus, all of these calculated values are also fuzzy numbers. Furthermore, if a crisp representation of the output is required, the center of area or other defuzzification methods can be employed. In the next section, the results of the proposed HFDES are compared with the analytically calculated results of simple queuing systems.

6.4 NUMERICAL EXPERIMENTS WITH SINGLE SERVER FUZZY QUEUEING SYSTEMS

The performance of the proposed HFDES methodology has been tested by already solved examples of fuzzy queuing systems in the literature. Examples of single server queuing systems with fuzzy membership functions (F) and exponential probability distributions (M) for inter-arrival times and service times are used (F/M/1, M/F/1). The HFDES is implemented in Simphony .NET (Hajjar and AbouRizk 2002). A simulation template called hybrid CYCLONE is developed in Simphony .NET. The elements of this template are the same as the ones developed for fuzzy CYCLONE in Chapter 4. The only difference is that in hybrid CYCLONE, COMBI and NORMAL elements accept both fuzzy numbers and probability distributions for their durations.

Two single server queueing examples are modelled with hybrid CYCLONE. Each simulation model is run for 10,000 entities passing through the simulation model. The queue performance measures of hybrid CYCLONE are calculated according to the proposed approach in Section 6.4. These performance measures are validated against analytically calculated queue performance measures. For the queueing systems that contain randomness and fuzziness, our experiments indicate that the hybrid CYCLONE produces results with the max absolute error of less than 3% for the upper and lower limits of the alpha-cuts. For example, consider the F/M/1 queueing system where the inter-arrival time is a trapezoidal fuzzy number equal to trap(4,6,7,8) and the service time is exponentially distributed with a mean equal to 2. The results obtained for average system length (L) with hybrid CYCLONE are compared with the results reported in Kao et al. (1999) (Table 6.1). The lower and upper limits of 11 alpha-cuts calculated by hybrid CYCLONE have an error of less than 3% relative to the analytical approach of Kao et al. (1999) as indicated in Table 6.1. The outputs are also represented in Figure 6.7 indicating that the result of hybrid CYCLONE is confirmed with the analytical result.

Table 6.1 Alpha-cuts for L for example of F/M/1 queueing example

alpha cut	Alpha cuts for *L* *HFR CYCLONE* Lower	Upper	Alpha cuts for *L* *Kao et al.* Lower	Upper
0	0.259316	0.641988	0.25506	0.6275
0.1	0.262894	0.591502	0.25857	0.57926
0.2	0.266576	0.549379	0.26219	0.53873
0.3	0.270365	0.513653	0.26593	0.50414
0.4	0.274271	0.482814	0.26978	0.47424
0.5	0.278313	0.455927	0.27376	0.44811
0.6	0.282501	0.432333	0.27787	0.42504
0.7	0.286823	0.411388	0.28212	0.40451
0.8	0.291295	0.392657	0.28651	0.38611
0.9	0.29592	0.375756	0.29106	0.3695
1	0.300709	0.360435	0.29578	0.35443

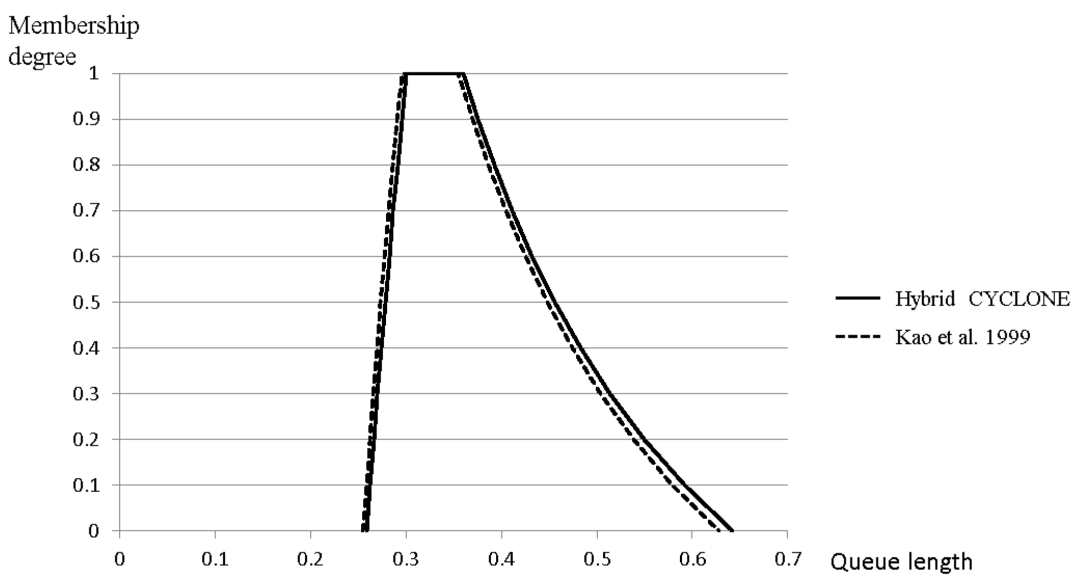


Figure 6.7 Comparison of the results of hybrid CYCLONE with analytically calculated results for average queue length

The performance measures of fuzzy queueing systems that are calculated with hybrid CYCLONE indicate complete compliance with the analytically solved performance measures in Kao et al. (1999). In the next section, the practicality of the proposed HFDES framework will be illustrated through a real case study of module assembly yard.

6.5 CASE STUDY OF MODULE ASSEMBLY YARD

The proposed HFDES is applied on a real case study of module assembly yard. Pre-assemble modules (e.g. piperack modules) are commonly used in building industrial construction projects (e.g. oil refinery plants) in northern regions of Alberta, Canada. Many of these pre-assembled modules are built in module assembly yards near Edmonton, AB and transported to the site with trailers. The current case study is developed for planning module assembly yard of a large construction company in Edmonton.

Because of the fast-track nature of the majority of industrial projects in Alberta, many projects performed in the module assembly yard are suffering from multiple change orders. These changes affect the productivity of different activities; thus have a cumulative effect on the cost and productivity in the module assembly yard. However, the impact of these change orders on the multi-project environment of the module assembly yard can be very complicated. Furthermore, the labour productivity in the module assembly yard is affected by other factors such as complexity of modules, weather conditions, and the skill level of the crew. Therefore, assessing the combined impact of various factors on the productivity of module assembly is very complicated.

The objective of this case study is to develop an integrated model for the module assembly yard of the partner company for estimating the impact of various factors on the productivity of individual activities and the overall production of the module assembly yard. This case study is conducted in the following 3 main steps:

1. Estimating the duration of activities that explicitly consider the impact of change orders and other significant factors that affect the productivity or production of the module assembly yard
2. Developing a HFDES template integrated with the prediction models (developed in step 1) for the module assembly process
3. Developing a simulation model for a module assembly project using the developed module assembly template (developed in step 2)

The above steps are discussed in Sections 6.5.1, 6.5.2, and 6.5.3 respectively.

6.5.1 Estimating Activity Durations

The module assembly process includes some typical activities (e.g. assembly structural steel) and some specific activities depending on type of a module. In this case study, 22 major activities that are included in many types of modules are identified for the module assembly process based on expert judgment (Table 6.2).

Table 6.2 Major activities in the module assembly process

Category	ID	Activity	Unit of measure	Number of Data Points
Structural Steel	1	Structural steel erection	Ton	971
	2	Grating	Meter^2	240
	3	Handrails/Kick Plates	Meter	190
	4	Structural welding	Inch	31
	5	Module building skid	Each	380
Piping	6	Handle pipes and fittings	Pound	166
	7	Rigs and install pipes	Meter	1410
	8	Fit and weld pipe spools	Each	709
	9	Valve and bolt-ups	Each	872
	10	Pipe supports/Gussets	Each	863
	11	Test and NDT	Meter	641
Insulation	12	Pipe insulation	Meter	206
	13	Equipment insulation	Meter^2	38
Electrical	14	Grounding	Each	128
	15	Cable Tray/Conduit	Meter	258
	16	Electrical Equipment	Each	81
	17	Electrical supports	Each	190
	18	Testing-Loop/Cont./Meg/	Each	48
Electrical Heat tracing	19	Tracing cable	Meter	549
	20	Power, End Kits/RTD	Each	337
	21	EHT Supports	Each	212
Painting	22	Painting	Meter^2	35

In order to estimate the duration of each activity, first, the productivity of the activity is estimated. The estimated productivity, p, is then used to estimate the activity duration, D, based on the number of labourers, l, shift hours per week, s, and quantity, q, using Equation 6.11. In this Equation, the unit of measurement of each of the variables are presented in parentheses in front of that variable. The unit for measuring the quantity of each of the activities is different as indicated in Table 6.2. Therefore, symbol u is used in Equation 6.11 to represent the measurement unit for quantity of activities.

$$D(week) = \frac{q(u)}{p\left(\frac{u}{man\_hour}\right) * s\left(\frac{hour}{week}\right) * l(man)} \tag{6.11}$$

It should be noted that, the estimated duration in Equation 6.11 represents the number of weeks that the workers are actively working on an activity and does not include the possible delays. The delays are later considered in the simulation model. Managers of the module yard decide about number of workers, l, and shift hours, s, based on different constraints such as availability of workers and due dates of modules. Furthermore, the quantity of each module activity, q, is a unique value that can be derived from the design documents of that module. On the other hand, the parameter p in the module assembly process is impacted by numerous factors which makes its estimation challenging. The methodology proposed in chapter 3 of this dissertation is used to predict the productivity of each of the activities in this case study. This methodology is based on the following steps: 1) Identifying factors and collecting data, 2) Developing a fuzzy rule-based system from data, 3) Estimating output uncertainty as a fuzzy number. The process of data collection of this case study has been discussed in chapter 3. Data are collected for 33 influencing factors on productivity of the activities in the module assembly yard.

Thus, 22 data sets (for 22 activities) with 33 inputs and 1 output (productivity) is collected. The number of data points for each activity is different ranging from 31 to 1410 (Table 6.2).

The feature selection for the inputs is performed using genetic algorithm as discussed in Chapter 3. Because the main objective in this case study is to predict the impact of factors related to change orders, some of the change order related factors are pre-set before applying GA for selecting features. For this purpose, the correlation coefficient of the factors under the category of change orders is estimated with the output. Two of these factors contained a significantly higher correlation coefficient compared with other factors, namely: 1) number of RFIs during execution of the activity, and 2) the number of revisions of the design documents. Thus, these two factors are pre-set before selecting the features using GA.

Based on the methodology proposed in chapter 3, fuzzy rule-based systems are developed for estimating the productivity of the activities listed in Table 6.2. The output of the fuzzy rule-based system is estimated as a triangular fuzzy number, as discussed in Chapter 3. Therefore, fuzzy arithmetic is used to estimate the activity duration as a fuzzy number according to Equation 6.11.

6.5.2 Integrated Simulation Template for Module Assembly Yard

Pipe spool modules have different designs and require different components. The modules are assembled using prefabricated components such as structural steel frames, cables, and pipe spool components that are fabricated in the pipe spool fabrication shop (Mohamed et al. 2007). Module assembly, where each module represents a unique project, is a complicated process. Each module also has unique features and requires a different number of resources as well as task sequences. Furthermore, modules share limited resources such as workers, space and equipment in the module assembly yard. The required materials for the modules should arrive to the module yard prior to starting each task. Moreover, each module has to be finished by specific due date indicated by the owner.

To start fabricating a module, first a proper spot in the module assembly yard has to be allocated to the module. The module yard consists of a number of bays in which the modules are assembled. According to the design of the modules, different tasks with different sequences are performed on each module. In the sequence of activities of a module, some of these tasks may appear more than once. For example, structural steel is usually performed at the start of building each level of a module, thus is repeated in each level. Furthermore, some of the tasks may overlap with each other; in other words, some portions of the tasks can be performed concurrently with other tasks. When all tasks of a module are finished, the module is shipped to the site using a truck (i.e. trailer). Dimensions of the module's envelop is designed so that the module can fit on the truck and be transported to the construction site (Davila Borrego 2004). Then the bay, assigned to a module, becomes available for other modules when the module is loaded in the truck.

Considering the above-mentioned differences among modules in the module assembly yard, a static scheduling model (e.g. CPM) cannot be developed to model different types of module. Therefore, a module yard template is developed in Simphony .NET based on the HFDES framework. This template allows modelling a module assembly project by connecting different activities in the form of a project network. The elements of the developed module assembly template are as follows:

- *Bay* element indicates the available number of bays in the module assembly yard.
- *Start Module* element creates the module entity at the specified point in time. The input of this element is the number of modules and start time of module assembly in the yard. This element generates a module entity in the simulation in the specified time if a bay is available to start the module. Otherwise, the element waits until a bay becomes available and the module will be started afterwards.
- *Activity* elements represent the activities in the module assembly yard. 22 elements are developed for the activities in Table 6.2. Each *activity* element is integrated with the fuzzy rule-based system that predicts the activity productivity as a fuzzy number (as discussed in Section 6.5.1). The inputs of each *activity* element are: delay duration (these delays can be due to various reasons such as the constraints in the project network of the module assembly process, late delivery of materials, weather conditions, etc.), man-hours assigned to the delay (man-hours that should be charged during the delay), number of workers assigned to that activity, quantity of the activity, working hours per week, and the input factors of the fuzzy-rule based system. Two of the factors related to change orders are pre-set in the input factors, namely: 1) number of RFIs (request for information) that happened during the process of the activity, and 2) the number of revisions of the design documents. However, other input factors are specific to the type of the activity. Two of the most common factors selected

for the activities are average skill level and one or more weather related factor such as temperature or precipitation. In the developed template, weather related factors are represented using stochastic uncertainty. For this purpose, the weather prediction component in Simphony .NET is employed (Wales and AbouRizk 1996). This component can generate temperature, precipitation and gust speed based on the statistical data of previous years. Furthermore, average skill level of crew is also modeled as a probability distribution based on the data of previous projects. On the other hand, the productivity estimated for the output of the fuzzy rule-based system is presented as a fuzzy number to represent the uncertainty of prediction. Figure 6.8 illustrates the inputs and components of *activity* elements.

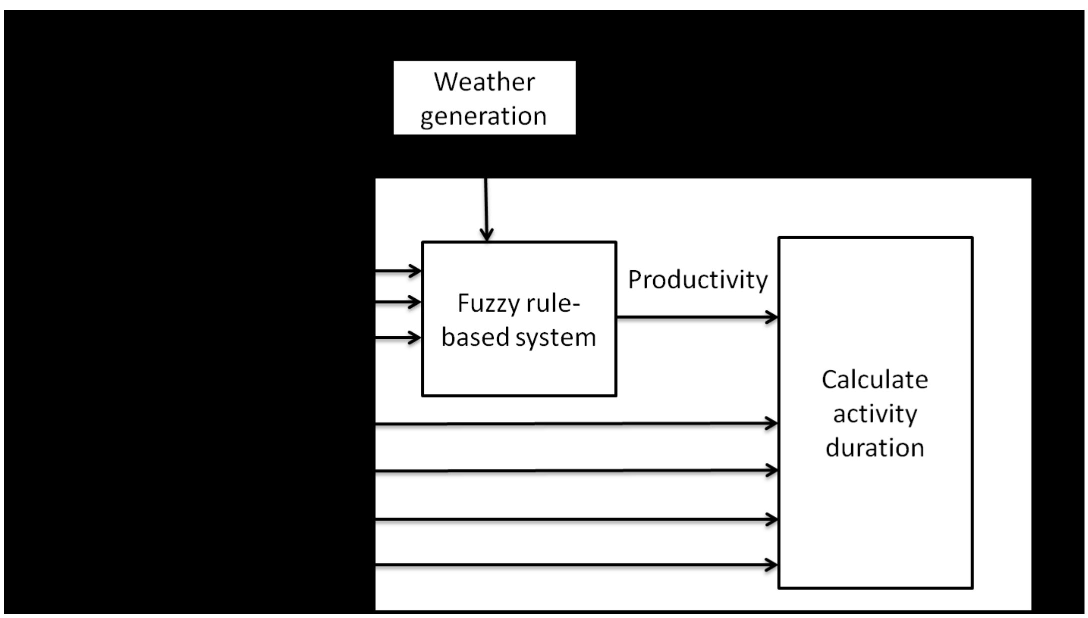


Figure 6.8 components of activity element in the module assembly template

- *Ship module* element indicates the end of the module assembly process and releases the bay captured by this module.

The simulation model provides outputs regarding the total duration of each module, man-hours spent for each module, and the total man-hours spent on the project. The simulation outcomes are sensitive to changes of the input factors such as the number of RFIs and skill level of labourers. The inputs of the simulation model can be represented with fuzzy numbers or probability distributions in HFDES. As discussed in this section, weather related factors and average crew skill level are modelled statistically. On the other hand, the productivity estimated for the output of the fuzzy rule-based system is presented as a fuzzy number to represent the uncertainty of prediction. Therefore, both fuzzy and stochastic uncertainties exist in this simulation template. In the next section, the discussed template is used to simulate a real case study of module assembly project.

6.5.3 Simulation Model of a Module Assembly Project

In this case study, one of the most recent projects of the company whose data has not been used in training the productivity prediction models is simulated. This project includes 30 modules with different start dates and minor differences in the sequence of their activities. Figure 6.9 represents the project network developed for one of the modules using the developed module assembly template. The complete project is modelled by developing a number of such project networks and assigning modules to those networks.

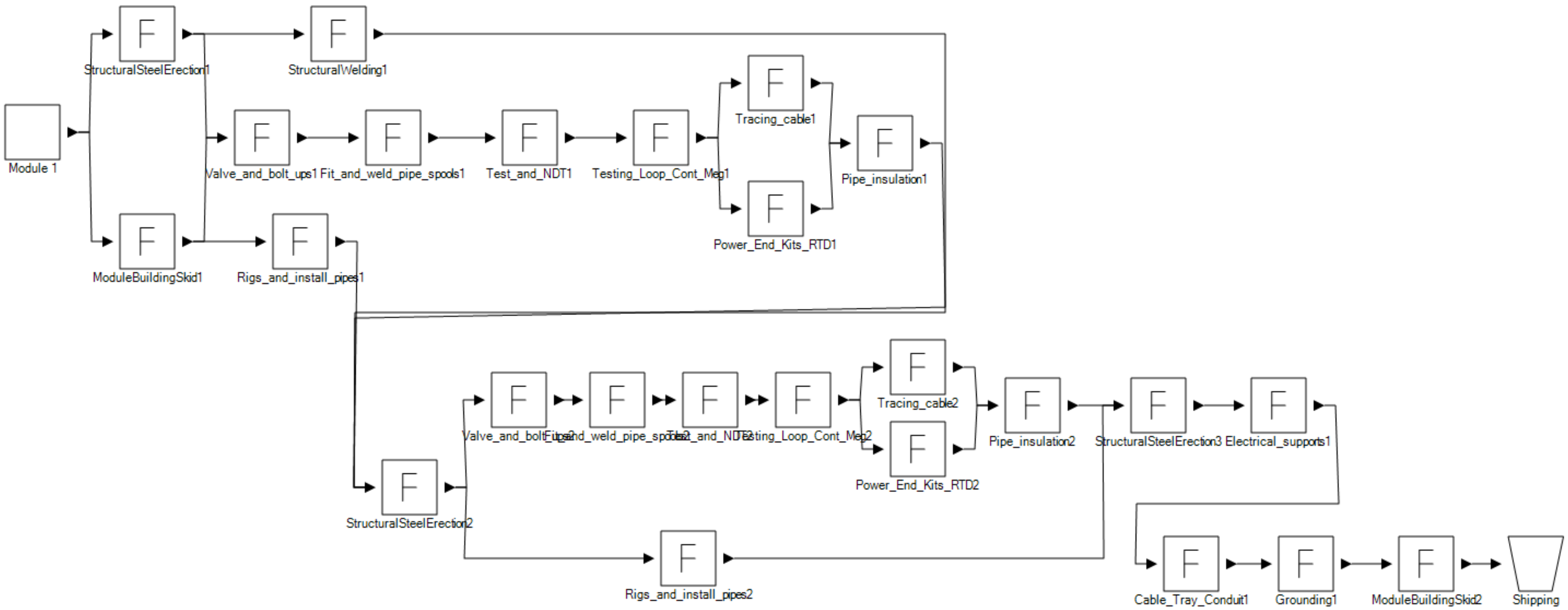


Figure 6.9 Simulation model for one of the modules in the module assembly yard

In the module assembly process simulated in Figure 6.9, once the required steel and spool pipes are mostly available (indicated by the start date of the module), the process of assembling a module starts in the yard. While part of structural steel is erected and the members are welded to make a stable structure, the process of building module skid and rig up, and erect pipe spools are performed in parallel. While valves are installed on the module are pipes are bolted up, electrical tasks such as tracing cables, testing loops, and placing electrical equipment and power kits are performed. After placing valves and fitting and welding pipe spools, the electrical NDT tests are performed and electrical supports are placed. Then the rest of steel structure is erected on the module, supports and gussets are placed and grating is performed. Afterwards, pipes are insulated and rest of electrical work, conduit, and placing cable trays, and grounding is performed. This process is accomplished by building module skid and wrap up the module.

The inputs to each of the activities are provided as crisp values, the required man-hours and project duration are calculated as the output of the simulation model. The developed simulation model can be used to experiment with the impact of modelled change order related factors: the number of RFIs and maximum revision numbers of module design documents. The developed model is sensitive to the changes in these two factors as they are considered as influencing factors in estimating all of the activity durations. By increasing or decreasing these factors for each of the modules, the impact in decreasing or increasing of the project duration or man-hours can be experimented.

As previously discussed, the developed simulation model contains both fuzzy (due to use of fuzzy rule-based system) and stochastic (due to random generation of weather conditions and average crew skill levels), thus HFDES framework is employed to process the uncertainties in the simulation. We have also developed models in the DES and FDES for the special cases where crisp (deterministic) values are used for weather conditions and output of fuzzy rule-based system, respectively. In HFDES weather related factors and average skill level of crew are modelled stochastically, while in FDES case, the mean value of the temperature, gust speed, and precipitation of the season in which the project took placed is used instead of stochastic generation of weather.

Figure 6.10 illustrates the CDF developed for the project completion time of the HFDES. The estimated completion of the project is compared with the completion time of same model developed with FDES and DES. In this figure, confidence index is used for the y-axis that is generalization of probability in the context of hybrid fuzzy and random environment as discussed in this chapter. In this figure, the range for the estimated duration is scaled to protect companies' confidential information.

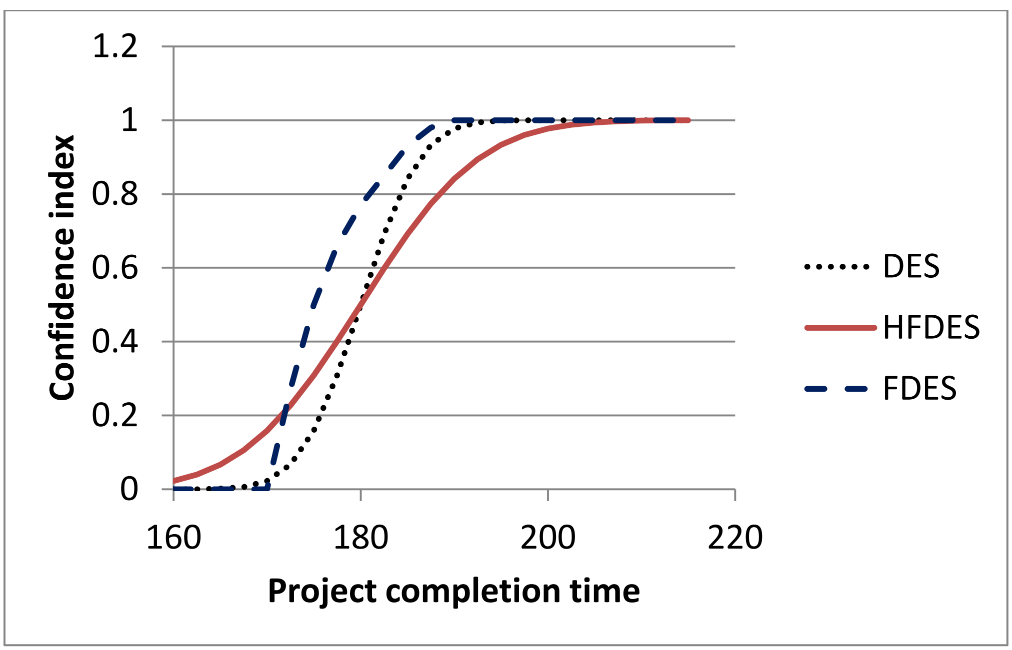


Figure 6.10 CDF of the project completion time estimated by DES, HFDES, DES

As it is illustrated in Figure 6.10, the range represented in the CDF estimated with HFDES is greater than the range estimated using DES or FDES. Furthermore, HFDES is slightly shifted toward right resulting in more conservative decisions. For example, the 90% quantile of the project completion time estimated by HFDES is 198 days. This means that one can be 90% confident that the duration of the project will not exceed 198 days. On the other hand, the 90% quantile of the project completion time estimated by FDES and DES are 182 and 187 days respectively. Thus HFDES estimates a more conservative value compared with DES and FDES, and is the only method that covers the actual project duration in 90% quantile. HFDES does not necessarily provide a more conservative estimate of the output, but provides a wider range of uncertainty. For example, for 20% confidence index, the project completion time is estimated as 165 days with HFDES, but is equal to 174 days with DES. Thus, depending on the chosen risk level, HFDES can be less or more conservative compared with DES. However, generally, HFDES is recommended in decision making of construction projects to account for subjective uncertainty. This is because more sources of uncertainties are considered in HFDES, which allows making more informed decisions.

Figure 6.11 represents the average project duration for HDES, FDES, and DES. The estimated project duration in this figure is scaled to protect the companies' confidential information. The actual project completion times is used as the benchmark for this case study. Traditional DES cannot be employed as a meaningful benchmark. This is because the results provided by DES only accounts for stochastic uncertainty and ignores the subjective uncertainty in the model. As illustrated in Figure 6.11, the average duration of the module assembly project using HFDES is closer to the actual project duration in the case study.

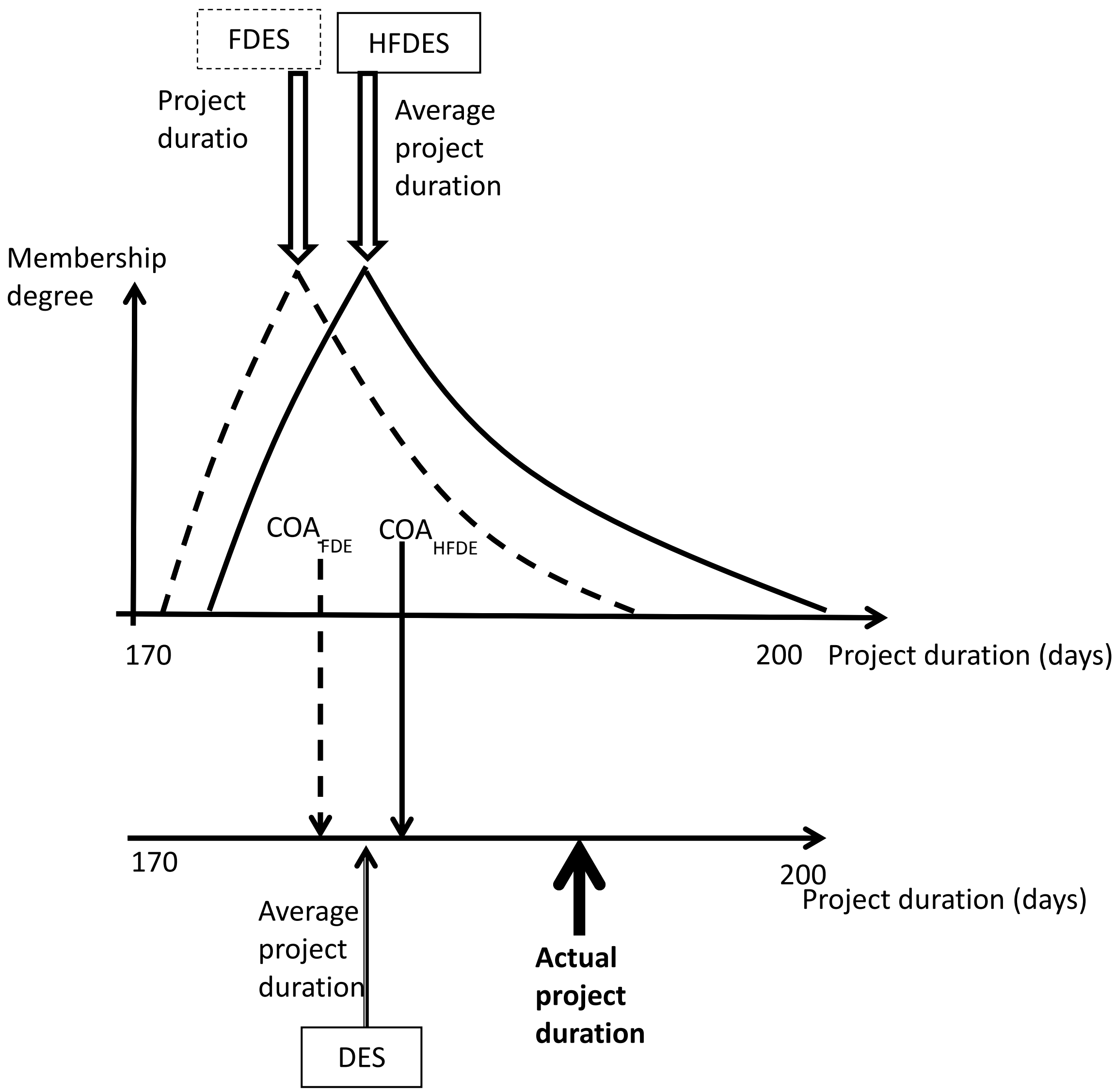


Figure 6.11 Comparison of the estimated average project completion time of a module assembly project with HFDES, FDES, DES and actual project duration

6.6 CONCLUDING REMARKS

A hybrid fuzzy discrete event simulation (HFDES) framework is proposed to simultaneously process both fuzzy and stochastic uncertainties. The proposed approach is based on sampling from stochastic uncertainties and employs FDES to process fuzzy uncertainties. A novel approach for calculating different statistical measures such as average queue length and waiting times in HFDES is also provided. The HFDES framework is validated using analytical results of queueing systems. Furthermore, a methodology is developed to provide a function for the output of HFDES which is equivalent to CDF in DES.

The proposed HFDES framework is implemented on a real case study of module assembly project to illustrate its practicality. In this case study, fuzzy rule-based systems are used for predicting productivity of the activities in the module assembly yard. A module assembly template, which is integrated with the fuzzy rule-based system, is developed for the module assembly process. The inputs of the fuzzy rule-based systems include features related to change orders along with other significant features. These features are selected using genetic algorithm. Thus, the simulation model developed with the module assembly template is able to evaluate the impact of change orders on the duration and required man-hours of the module assembly projects. The developed fuzzy rule-based systems integrated with the module assembly template contain one or more weather related inputs such as temperature, wind speed, or precipitation. For these inputs, as well as, average crew skill level, probability distributions derived from historical data. On the other hand, fuzzy numbers are used to represent the uncertainty of the output of the fuzzy rule-based systems. Thus, both fuzzy and stochastic uncertainties are encountered in this case study requiring a HFDES. The developed module assembly template in the HFDES is used to estimate the productivity and required man-hours of one of the projects in the module assembly yard. The results of the HFDES for this project were closer to the actual duration and man-hours of the module assembly yard in comparison with DES or FDES models of the same project.

6.7 REFERENCES

Baudrit, C., Dubois, D., & Perrot, N. (2008). Representing parametric probabilistic models tainted with imprecision. *Fuzzy sets and systems*, *159*(15), 1913-1928.

Kao, C., Li, C. C., & Chen, S. P. (1999). Parametric programming to the analysis of fuzzy queues. Fuzzy Sets and Systems, 107(1), 93-100.

Chen, Z., Huang, G. H., & Chakma, A. (2003). Hybrid fuzzy-stochastic modeling approach for assessing environmental risks at contaminated groundwater systems. *Journal of Environmental Engineering*, *129*(1), 79-88.

Chen, Z., Zhao, L., & Lee, K. (2010). Environmental risk assessment of offshore produced water discharges using a hybrid fuzzy-stochastic modeling approach. *Environmental Modelling & Software*, *25*(6), 782-792.

Cooper, J. A., Ferson, S., & Ginzburg, L. (1996). Hybrid processing of stochastic and subjective uncertainty data. *Risk Analysis*, *16*(6), 785-791.

Davis, T. J., & Keller, C. P. (1997). Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. *International Journal of Geographical Information Science*, *11*(5), 409-434.

Dubois, D., & Guyonnet, D. (2011). Risk-informed decision-making in the presence of epistemic uncertainty. *International Journal of General Systems, 40(02*), 145-167.

Hajjar, D., & AbouRizk, S. (1999). Simphony: an environment for building special purpose construction simulation tools. *Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future-Volume 2,* pp. 998-1006, ACM.

Huang, G. H. (1998). A hybrid inexact-stochastic water management model.*European Journal of Operational Research*, *107*(1), 137-158.

Kentel, E., & Aral, M. M. (2005). 2D Monte Carlo versus 2D fuzzy Monte Carlo health risk assessment. *Stochastic Environmental Research and Risk Assessment*, *19*(1), 86-96.

Liu, L., Huang, G. H., Liu, Y., Fuller, G. A., & Zeng, G. M. (2003). A fuzzy-stochastic robust programming model for regional air quality management under uncertainty. *Engineering Optimization*, *35*(2), 177-199.

Möller, B., & Beer, M. (2004). *Fuzzy randomness: uncertainty in civil engineering and computational mechanics*. Springer.

Sadeghi, N., Fayek, A. R., & Pedrycz, W. (2010). Fuzzy Monte Carlo simulation and risk assessment in construction. *Computer-Aided Civil and Infrastructure Engineering*, *25*(4), 238-252.

Wales, R. J., & AbouRizk, S. M. (1996). An integrated simulation model for construction. Simulation practice and theory, 3(6), 401-420.

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

A review of the work conducted in this research, and a summary of the contributions is provided. Furthermore, limitations of the developed model and recommendations for future research are outlined.

7.1 RESEARCH SUMMARY

Discrete event simulation (DES) has been previously employed for planning and analyzing construction projects. One of the main advantageous of using DES in construction management is its capability in considering uncertainty of different variables in construction processes. Traditionally, probability distributions (i.e. stochastic uncertainty) are used to represent uncertainties in DES. However, uncertainty can be categorized as subjective and stochastic. Stochastic uncertainty is associated with actual variation of a variable. On the other hand, subjective uncertainty represents imprecision and lack of knowledge. Subjective uncertainty is often encountered in construction simulation due to the lack of data, linguistic expression and use of expert judgment in estimating activity durations. Fuzzy set theory provides a methodology for mathematical modelling of subjective uncertainty. However, traditional DES frameworks are not able to consider fuzzy inputs to the simulation model.

Recently, fuzzy discrete event simulation (FDES) has been proposed for construction management as an integration of fuzzy set theory and DES for dealing with subjective uncertainty in construction simulation. Although FDES can greatly benefit the simulation of construction management applications, the fundamental differences between fuzzy numbers and probability distributions introduce new challenges to FDES. These challenges have to be addressed for effective usage of FDES in construction management.

Furthermore, Fuzzy and stochastic uncertainties are complementary methods of representing uncertainty and can simultaneously exist in a simulation model. For example, some of the activity durations may be represented using fuzzy numbers due to the lack of data and linguistic expression of knowledge, while other input parameters may be represented using probability distributions that are developed from data. However, Traditional DES can only handle stochastic uncertainty (represented by probability distributions) and FDES can only handle subjective uncertainty (represented by fuzzy numbers); A DES framework that can handle both types of subjective and stochastic uncertainties was not proposed in previous literature.

In this research, we developed a framework for DES of construction projects that is able to handle both subjective and stochastic uncertainty in activity durations of construction projects. The gaps that are identified in the literature, the proposed methodologies and frameworks to address those gaps, as well as the validation approaches that are employed to validate these methodologies are summarized in Table 7.1.

Table 7.1 Gaps in the literature, proposed approaches in this dissertation to address those gapes, and employed validation methodologies

Gaps in previous research	Proposed methodologies to address gaps	Validation	Chapter
Unavailability of a framework to develop interpretable data-driven productivity prediction models and accounting for subjective uncertainty in these models	A fuzzy rule-based data-driven framework for developing interpretable activity duration prediction models	Estimating the productivity of activities of an actual case study of module assembly yard	Chapter 3
Sensitivity of the outcomes of FDES to fuzzy ranking methodology and the problem of time paradox	A new approach for advancing the simulation time based on the logical dependencies of the event times	Project network of building construction and tunnelling case study	Chapter 4
Unavailability of a methodology for analysis of queues in FDES	A new approach for analysis of queues in FDES by developing a correlation network	Fuzzy queueing examples from the literature and the example of earthmoving operation	Chapter 5
Unavailability of a framework for considering both subjective and stochastic uncertainties in DES	A HFDES framework for considering both subjective and stochastic uncertainties	Fuzzy queueing examples from the literature and case study of module assembly yard	Chapter 6

This research is conducted in three main stages: 1) enhancing approaches for estimating the uncertainties of construction activities, 2) enhancing the state of the art of FDES for construction management; 3) developing a hybrid fuzzy discrete event simulation (HFDES) framework that can handle both fuzzy and stochastic uncertainties.

7.1.1 The First Stage

A brief review of methods for estimating durations and productivities of construction activities is provided in the second chapter. These methods are discussed in two categories:

- In the first category, durations or productivities of activities are estimated using probability distributions or fuzzy numbers. The impact of different factors on productivity or durations of construction activities are implicitly considered in these estimates.
- In the second category, a prediction model is developed that can explicitly consider the impact of different factors on durations or productivities of construction activities. These prediction models may be developed from data or by using expert knowledge.

In either case, the values estimated using a prediction model contain uncertainties due to the inaccuracy of prediction. These uncertainties are sometimes large in models that are predicting durations or productivities of construction activities. However, very limited effort in the area of construction engineering and management has been made for representing these uncertainties. Generally the uncertainties of prediction models may be presented using probability distributions or fuzzy numbers depending on the availability of data and type of prediction model.

In Chapter 3, a comprehensive framework for developing interpretable construction productivity prediction models based on fuzzy rule-based systems is proposed. The proposed approach uses fuzzy C-means clustering to develop an initial rule-based system. Furthermore, genetic algorithm is used to optimize the features and parameters of the fuzzy rule-based system. A novel approach has been also developed to model the output of the fuzzy rule-based system as a fuzzy number to represent the uncertainty of prediction.

Representing the output uncertainty of the fuzzy rule-based system as proposed in Chapter 3 is especially important when performing further analysis with the predicted productivity. For example, the estimated fuzzy productivity may be used to estimate the activity duration as a fuzzy number. This activity duration can become the input to a project network or a simulation model which will be then used for if then-analysis, optimization, or scheduling.

7.1.2 The Second Stage

For DES of construction projects with fuzzy activity durations, a FDES framework is required. Two main shortcomings of previous FDES frameworks are identified and approaches proposed to overcome these shortcomings in chapters 4 and 5.

- Chapter 4 illustrates that the available approaches of fuzzy discrete event simulation (FDES) either overestimate or underestimate the simulation time. A new approach for calculating the event times in FDES to eliminate these shortcomings was proposed. The proposed approach was tested with different structures of CYCLONE modelling elements and was shown to eliminate the problem of overestimation or underestimation of the simulation time. Additionally, the proposed approach of FDES was implemented. A project network of building construction is used to show that the results of the proposed FDES methodology produce the same results as the analytically-calculated results. Furthermore, an actual case study of a tunnelling construction operation is also presented in Chapter 4. It is demonstrated that by representing the fuzziness in the estimated penetration rate of the TBM machine, the simulation output is able to represent the subjective uncertainty caused by the use of linguistic expression of knowledge.

- Chapter 5 acknowledge the importance of calculating queue performance measures in construction management. However, previously developed FDES frameworks only calculate simulation time (e.g., project completion time) for the simulation output; they do not have the capability of calculating queue performance measures such as average queue length and waiting time. Chapter 5 proposes an approach for calculating average queue length and waiting time in FDES. First, a correlation network is developed to track the correlation of event times. The subtraction of fuzzy event times for calculation of average queue length and waiting time is then performed by considering the developed correlation network. The proposed approach is validated using analytically solved queueing examples. Furthermore, the practicality of the proposed approach is illustrated using an example of asphalt paving operation in which the number of trucks has been optimized.

Thus, in stage 2, a FDES framework is developed for processing fuzzy activity durations. However, this framework is not capable in dealing with fuzzy and stochastic uncertainties simultaneously. In the next stage, a framework is proposed that enables simultaneous processing of fuzzy and stochastic uncertainties.

7.1.3 The Third Stage

In Chapter 6, a hybrid fuzzy discrete event simulation (HFDES) framework is proposed that can simultaneously process fuzzy and stochastic uncertainties. The proposed approach is based on sampling from stochastic uncertainties and employs FDES to process fuzzy uncertainties. Furthermore, a methodology is developed to provide a function for the output of HFDES which is equivalent to cumulative distribution function in DES. The approach for calculating different statistical measures such as average queue length and waiting times in HFDES is also discussed. The HFDES framework is validated using analytical results of queueing systems.

The proposed FDES framework is implemented on a real case study of module assembly process. In this case study, fuzzy rule-based systems are used for predicting the productivity of the activities in the module assembly yard. These fuzzy rule-based systems are developed based on the methodology proposed in Chapter 3. A module assembly template is developed for the module assembly process that is integrated with the fuzzy rule-based systems. The inputs of the fuzzy rule-based systems include features related to change orders along with other significant features that are selected using genetic algorithm. Thus, simulation models developed with the module assembly template can evaluate the impact of change orders on the duration and required man-hours of the module assembly projects. All of the developed fuzzy rule-based systems integrated with the module assembly template contain weather related inputs such as temperature, wind speed, or precipitation, as well as, average crew skill level. For these inputs, input values are generated stochastically based on historical data. On the other hand, fuzzy numbers are used to represent the uncertainty of the predicted results. Thus, both fuzzy and stochastic uncertainty is encountered in this case study requiring a HFDES. The developed module assembly template in the HFDES is used to estimate the productivity and required man-hours of a project in the module assembly yard. The results of the HFDES for this project were closer to the actual duration and man-hours of the module assembly yard in comparison with DES or FDES models of the same project.

7.2 RESEARCH CONTRIBUTIONS

This research study presents various academic and industrial contributions to the construction industry. The details of these contributions are as follow.

7.2.1 Academic Contributions

The main academic contributions offered by this research are summarized as follows:

1. *Development of a novel framework to predict construction productivity based on interpretable fuzzy rule-based systems:* The proposed framework explicitly models the impact of different factors on productivity. The interpretability of the developed framework is the main advantage of the proposed framework compared with artificial neural networks which are the most common approaches for developing data-driven productivity prediction models in recent years. Furthermore, the input factors of the proposed fuzzy-based productivity prediction model are selected using a feature selection approach based on genetic algorithm for the first time.
2. *Development of a methodology for considering output uncertainty of a data-driven fuzzy rule-based system as a fuzzy number:* In previous construction productivity prediction models, the development of probability distribution or fuzzy number to represent the uncertainty of prediction is often ignored. The proposed approach enables representing the output uncertainty of the fuzzy rule-based system as a triangular fuzzy number. This fuzzy number is

estimated based on the theory of justifiable granularity and genetic algorithms to represent the uncertainty of prediction.

3. *Advancement of previous fuzzy discrete event simulation (FDES) frameworks for considering subjective uncertainty in construction projects:* This framework eliminates the problem of overestimation and underestimation of event times in the FDES, which was a main issue in previously proposed FDES frameworks. This framework is based on a new methodology for advancing the simulation time in FDES that considers the logical relationships of the simulation events.
4. *Development of a methodology for analysis of queues in FDES:* The proposed methodology extends the capability of previously developed FDES frameworks by providing an approach for analysis of queues. Generally, subtraction of event times is required for calculating queue performance measures such as waiting time. Discussion in Chapter 5 indicates that the subtraction of fuzzy event times in FDES cannot be easily performed as the event times are correlated to each other. In the proposed approach, a correlation network is developed to track the correlation of event times. The subtraction of fuzzy event times for calculation of average queue length and waiting time is then performed by considering the developed correlation network.
5. *Introduction of a hybrid fuzzy discrete event simulation (HFDES) framework that can simultaneously process both subjective and stochastic uncertainties in discrete event simulation:* In the proposed methodology, stochastic input

parameters are modelled with probability distributions and subjective input parameters are presented as fuzzy numbers. Previous event-based simulation frameworks can either consider probability distributions or fuzzy numbers as the simulation inputs. In other words, a framework that can model both types of uncertainties was not previously available. Thus, the proposed approach in this dissertation advances the state of the art by allowing the modeller to represent both types of uncertainties simultaneously.

7.2.2 Industrial Contributions

In addition to the academic contributions, this research also offers several industrial contributions, from which companies involved in construction and industrial projects can benefit. These contributions are summarized as follows:

1. *Facilitation of developing simulation models in construction industry*: One of the main challenges of applying simulation model in construction projects is collecting reliable input data to develop input probability distributions. This is because enough historical data are not usually available for some of the input variables of construction simulation models. The proposed framework of HFDES enables developing fuzzy numbers based on expert judgement without having enough historical data available. At the same time, we can employ probability distributions for inputs for which data are available. Thus, using HFDES, unavailability of data will not be a limiting factor for developing simulation models of construction projects. Having said that, the estimations of simulation inputs based on accurate historical data are often

more accurate compared with the estimations of simulation inputs based on expert judgment. On the other hand, HFDES allows estimators/planners to consider the impact of the possible uncertainty (i.e. imprecision) due to the use expert knowledge.

2. *Consideration of the impact of subjective uncertainties as well as stochastic uncertainties in simulation outputs*: Ignoring some sources of uncertainty (either subjective or stochastic) may happen in DES or FDES simulation models because of the lack of capability of capturing subjective or stochastic uncertainties, respectively. The ignorance of one source of uncertainty would result in unrealistically precise simulation outcomes. This dissertation proposes a hybrid HFDES framework that can represents both subjective and stochastic uncertainties. Thus, the proposed HFDES framework enables construction companies to make more informed decisions from simulation outputs.
3. *Development of a framework based on fuzzy rule-based systems for predicting the productivity or duration of project activities*: This framework provides an interpretable fuzzy rule-based system to explicitly model the impact of different factors on the productivity of a project (e.g. module assembly construction). The proposed framework enables construction companies to perform if-then analysis by changing the input parameters of a model and estimating the impact of different parameters on the productivity of individual activities.

4. *Development of an integrated simulation framework to consider the impact of various uncertain factors on productivity and duration of industrial construction projects*: In the proposed framework, fuzzy rule-based systems are integrated with the simulation model to estimate the activity durations of construction projects based on different factors. The proposed framework enables construction companies to analyze the impact of different factors on the overall duration and to estimate the resource requirements of industrial construction projects.

7.3 RESEARCH LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH AND DEVELOPMENT

This research provides a basis for simultaneous consideration of subjective and stochastic uncertainties in discrete event simulation of construction projects. Despite the contributions presented in this research, the research has some limitations that are recommended to be addressed in future research projects. The limitations and recommendations for areas of enhancements are provided in the following subsections.

7.3.1 Productivity Prediction Model

The proposed framework of developing productivity prediction model contains some limitations that can be enhanced in the future in the following aspects:

- The proposed approach of developing the fuzzy rule-based system employs a genetic algorithm to optimize the input features and parameters of membership functions. However, a fuzzy rule-based system includes other parameters that may be optimized for optimum performance. In the future, other parameters of the fuzzy rule-based system such as the number of clusters, the weight of the rules, and the defuzzification methodology can be also subjected to optimization.

- The sensitivity of the simulation outcomes to the parameters of the fuzzy membership functions defining the input should be further investigated in the future. For example, the impact on the simulation output of the opinion of different experts on the input membership functions should be investigated.
- The proposed fuzzy-rule based approach is classified as genetic-fuzzy model. The performance of other methods of developing interpretable prediction models (e.g. neuro-fuzzy systems) should be investigated in the future.
- One of the possible advantageous of developing interpretable fuzzy rule-based systems is the possibility of later modification based on expert judgment. Future research is recommended on updating the fuzzy rule-based system based on expert judgment. These updates can be, for example, due to changes in context variables such as management policies or location.

7.3.2 Fuzzy Uncertainty of Simulation Parameters

FDES enable considering subjective uncertainty in activity durations of construction projects. The sensitivity of the developed fuzzy rule-based system to the changes of input factors is recommended for further investigation in the future. Specifically, the sensitivity and accuracy of the fuzzy rule-based system compared to other types of productivity prediction models developed with the same data, such as artificial neural networks, is recommended.

Furthermore, although activity durations are one of the most uncertain factors in construction simulation models, other input parameters of construction simulation models may also contain subjective uncertainty. For example, the number of available resources may be uncertain and expressed by experts as a fuzzy number. However, current FDES frameworks are not able to process fuzzy uncertainty in the total number of resources. It is recommended that in the future research, FDES is extended to cover this aspect.

7.3.3 Imprecise Probability Distributions

The proposed HFDES framework is capable of modelling both fuzzy and stochastic uncertainties. However, as discussed in chapter 6, when estimating the input parameters of the simulation model, some of the simulation parameters may have a random nature. When data are not available for those parameters and the type of probability distribution is not known, fuzzy numbers can be used to represent the uncertainty of those parameters. On the other hand, if the type of probability distribution is known, expert may estimate a probability distribution for those parameters. However, the parameters of probability distributions estimated using this approach are often not precise and may be represented with an interval or a fuzzy number. A probability distribution of which its parameters are defined as an interval or a fuzzy number represents an imprecise probability distribution. A simulation framework that can handle imprecise probabilities in its input parameters is recommended to be developed in the future.

7.3.4 Analysis of Queues

The analysis of queues containing subjective uncertainty is provided as an extension to FDES in this dissertation. The proposed approach is based on developing a correlation network of the event times in FDES. However, although, the approach provides correct results, as validated using analytically solved queueing examples, it is not computationally efficient for big simulation models involving millions of events. This is because in big simulation models, the proposed correlation network is very big, thus searching through the network to find possible correlations between event times is very time consuming. As discussed in the thesis, the number of generations that are tracked in the correlation network is limited to increase the speed of the simulation model. In the future, further analysis of the implications of limiting the number of generations on the accuracy of the calculated queue performance measures is recommended. Furthermore, it is recommended to investigate possible improvements such as more efficient searching through the correlation network to improve this efficiency.

7.3.5 Fuzzy Discrete Event Simulation Framework

The proposed FDES framework minimizes the impact of fuzzy ranking and solves the issue of overestimation and underestimation of simulation time. However, as discussed in Chapter 4, the proposed FDES framework still has the limitation of considering only one possible path of the entities in the simulation. For example, when a queue is followed by a number of activities, the activity that will be chosen to process the entity depends on the ranking order of the event times. developed fuzzy rule-based system to the changes of input.

7.3.6 Integrated Simulation Template of Module Assembly Yard

The integrated module assembly yard template enables performing if-then analysis for the impact of modelled input factors on the durations and required man-hours of the projects in the module assembly yard. However, this template is currently developed based on the data collected from one company. Therefore, the developed model is specific to certain context variables such as location and management policies of the company. In the future, further data should be collected from other companies in order to develop a generalized template for simulating module assembly projects.

7.3.7 Applications of FDES and HFDES in Scheduling

In this thesis, FDES and HFDES are employed for if-then analysis and estimating different simulation outputs such as project completing time and waiting time. On the other hand, simulation has been previously used for schedule optimization of construction projects. FDES or HFDES may be employed for managing uncertainties in scheduling of construction projects in the future. For example, some preliminary works have been provided by this researcher for proactive scheduling with fuzzy activity durations as indicated in Appendix B.

APPENDIX A: SURVEY FOR MODIFYING THE LIST OF FACTORS FOR MODULE ASSEMBLY YARD

Factor/Sub-factor		Measuring Method	Impact					Data Availability					Alternative Measuring Method
			1. Very Low/ No impact	2. Low	3. Moderate	4.High	5. Very High	1. very Low/ No data	2. Low	3. Moderate	4.High	5. Very High	
Project Manager			1	2	3	4	5						
	Historical productivity	Productivity (Real number)	1	2	3	4	5	1	2	3	4	5	
	Skill level	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Superintendent			1	2	3	4	5						
	historical performance	Productivity (Real number)	1	2	3	4	5	1	2	3	4	5	
	Capability to organize activities	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Client			1	2	3	4	5						
	Timely response to questions and inquiries	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Average time to respond to questions and inquiries	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Quality of coordination with client	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Years with client	#Years (Real number)	1	2	3	4	5	1	2	3	4	5	

Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Consultants			1	2	3	4	5						
	Average time for verifying progress claims	#Days ()	1	2	3	4	5	1	2	3	4	5	
	Quality of coordination with consultants	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Subcontractor			1	2	3	4	5						
	Timely finishing of tasks by sub-contractors	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Average amount of delays by sub-contractors	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Quality of coordination with subcontractors	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Contract			1	2	3	4	5						
	Type of contract	Guaranteed maxi mum price/ Lump sum/ Unit price/ Cost plus/ Cost-reimbursable alternative/ Integrated project delivery or alliance (Categorical)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Economy			1	2	3	4	5						

	Economy Conditions	Very poor/Poor/Moderate/Good/Very good	1	2	3	4	5	1	2	3	4	5	
	Inflation rates	%Rate	1	2	3	4	5	1	2	3	4	5	
	Bank interest rates	%Rate	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Safety			1	2	3	4	5						
	Recordable incident rate(RIR)	%Rate	1	2	3	4	5	1	2	3	4	5	
	Lost workday case incident rate(LIR)	%Rate	1	2	3	4	5	1	2	3	4	5	
	Having adequate safety plans and practices for the project	Strongly agree/Agree/Natural/Disagree/Strongly disagree	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
2. Project Difficulty													
Complexity			1	2	3	4	5						
	Overall project complexity	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Past experience with construction methods	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Execution plan	Stick built steel/ Assembled frames (Categorical)	1	2	3	4	5	1	2	3	4	5	

	Estimated man-hours for the project	#Man-hours (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Change order			1	2	3	4	5						
	Number of change orders in the project	#Change order (Real number)	1	2	3	4	5	1	2	3	4	5	
	Timing of change	#Changes occurred after job is 50% complete (Real number)	1	2	3	4	5	1	2	3	4	5	
	Processing time for change orders	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Ratio of approved total volume of change order to total work volume	%Man-hour (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Drawing and Design			1	2	3	4	5						
	Completeness of Drawings	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Readability and Clarity of Drawings	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Buildability (constructability) of design	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	

	Number of drawing revisions	#Revisions (Real number)	1	2	3	4	5	1	2	3	4	5	
	Experience of design team	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	waiting time for approval of drawings	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Early Availability of Drawings	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Unforeseen factors			1	2	3	4	5						
	Flood/ Earthquake/ Fire	Flood/ Earthquake/ Fire (Categorical)	1	2	3	4	5	1	2	3	4	5	
	Workers' strikes	Number of days on strike (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Extra work			1	2	3	4	5						
	Percentage of man-hours spend on extra work that were beyond the original scope of work	%Man-hours (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Rework			1	2	3	4	5						

	Construction filed rework index	%Ratio of Activity total Cost of rework to total field construction phase cost (Real number)	1	2	3	4	5	1	2	3	4	5	
	Frequency of rework	Number of rework occurrence per scope of work (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Schedule			1	2	3	4	5						
	Quality of project schedule	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Unrealistic deadline set by client	Number days that the project duration is underestimated (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Financial problems			1	2	3	4	5						
	Delay in issuing payments by client	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Existence of financial problems	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
3. General Activity													
Learning			1	2	3	4	5						

	Number of similar modules in the project	#Similar modules (Real number)	1	2	3	4	5	1	2	3	4	5	
	Past experience with configuration and geometry	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
General module specifications			1	2	3	4	5						
	Height	Ft. (Real number)	1	2	3	4	5	1	2	3	4	5	
	Length	Ft. (Real number)	1	2	3	4	5	1	2	3	4	5	
	Width	Ft. (Real number)	1	2	3	4	5	1	2	3	4	5	
	Weight	Ton (Real number)	1	2	3	4	5	1	2	3	4	5	
	Number of levels	#Levels (Real number)	1	2	3	4	5	1	2	3	4	5	
	Type	(Categorical)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
4. Activity Difficulty													
Complexity of activity			1	2	3	4	5						
	Technological complexity of activity	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Overall activity difficulty	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Delays and interruptions			1	2	3	4	5						

	Number of Interruptions of an activity	#Interruptions (Real number)	1	2	3	4	5	1	2	3	4	5	
	Duration of each Interruption	#Hours (Real number)	1	2	3	4	5	1	2	3	4	5	
	Delay of the activity	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Duration of each Interruption	#Hours (Real number)	1	2	3	4	5	1	2	3	4	5	
	Delay of the activity	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	#Difference between the start time of the activity in the initial and actual schedule	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	
	Out of sequence work	Percentage of work performed out of sequence (Real number)	1	2	3	4	5	1	2	3	4	5	
	Increase in man-hours from the original schedule	#Man-hours (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Timing of task			1	2	3	4	5						
	Night shift man-hours/total budgeted man-hours	%Man-hours (Real number)											

	Weekend and holiday man-hours/total budgeted man-hours	%Man-hours (Real number)	1	2	3	4	5	1	2	3	4	5	
	Overtime man-hours/total budgeted man-hours	%Man-hours (Real number)	1	2	3	4	5	1	2	3	4	5	
Weather			1	2	3	4	5						
	Season	Spring/Sumer/Fall/ Winter (Categorical)	1	2	3	4	5	1	2	3	4	5	
	Wind	Km/hour (Real number)	1	2	3	4	5	1	2	3	4	5	
	Humidity	%Humidity (Real number)	1	2	3	4	5	1	2	3	4	5	
	Temperature	Degree centigrade (Real number)	1	2	3	4	5	1	2	3	4	5	
	Precipitation	No Precipitation/Drizzle and Flurries/ Rain / Snow (Categorical)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Congestion			1	2	3	4	5						
	Stacking of trades	#Concurrent activities on the module (Real number)	1	2	3	4	5	1	2	3	4	5	
	Ratio of peak man power to average man power	Ratio (Real number)	1	2	3	4	5	1	2	3	4	5	
	Number of people per square feet	#people/ft2 (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Crew			1	2	3	4	5						

	Average crew size	#Crew (Real number)	1	2	3	4	5	1	2	3	4	5	
	Peak crew size	#Crew (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Inspection and quality			1	2	3	4	5						
	High inspection requirements	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Response of consultant staff to attend inspection work	#Days of delay to inspect (Real number)	1	2	3	4	5	1	2	3	4	5	
	High quality requirements	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Safety of working condition			1	2	3	4	5						
	Risk level of the activity	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Cost of accidents	$ (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Distance to facilities			1	2	3	4	5						
	Distance to material storage	Feet (Real number)	1	2	3	4	5	1	2	3	4	5	
	Distance to lunch area	Feet (Real number)	1	2	3	4	5	1	2	3	4	5	
	Distance to washroom	Feet (Real number)	1	2	3	4	5	1	2	3	4	5	

Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
5. Foremen and Tradesmen													
Skill of labour			1	2	3	4	5						
	Labours skill level	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Unionized	Yes/ No (Boolean)	1	2	3	4	5	1	2	3	4	5	
	Foreign or local	Foreign/ Local (Categorical)	1	2	3	4	5	1	2	3	4	5	
	Multi-tasking	Yes/No (Boolean)	1	2	3	4	5	1	2	3	4	5	
	Appropriate training of labour	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Worker motivation			1	2	3	4	5						
	Morale	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Labour trust in supervision	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Skill of foreman			1	2	3	4	5						
	Adequacy of instructions	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Foreman training	#Attended trainings*duration of training (Real number)	1	2	3	4	5	1	2	3	4	5	
	Foreman historical performance	Productivity (Real number)	1	2	3	4	5	1	2	3	4	5	
	Foreman years of experience	Real number (Real number)	1	2	3	4	5	1	2	3	4	5	

	Appropriate arrangement of crew by foreman	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Labours availability			1	2	3	4	5						
	Availability of skilled workers	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Absenteeism	#Absent man-hours over total man-hours= %Absenteeism (Real number)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
6. Equipment and Material													
Equipment			1	2	3	4	5						
	Equipment cost per direct man-hour	$ (Real number)	1	2	3	4	5	1	2	3	4	5	
	Availability of equipment	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Quality of equipment	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	
Material			1	2	3	4	5						
	Material cost per direct man-hour	$ (Real number)	1	2	3	4	5	1	2	3	4	5	
	Delay in arrival of materials	#Days (Real number)	1	2	3	4	5	1	2	3	4	5	

	Experience with the material type	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
	Availability of material	Scale 1 to 7 (Qualitative rating)	1	2	3	4	5	1	2	3	4	5	
Other Sub-factors			1	2	3	4	5	1	2	3	4	5	

...

APPENDIX B: A FUZZY-BASED APPROACH FOR PROACTIVE SCHEDULING OF CONSTRUCTION PROJECTS[3]

Naimeh Sadeghi[1], and Aminah Robinson Fayek[2]
[1, 2] Department of Civil and Environmental Engineering, Hole School of Construction, University of Alberta, Edmonton, Alberta, Canada T6G 2W2

Abstract: Construction projects are usually subject to a high degree of uncertainty. Most traditional approaches for developing a construction project schedule are based on considering complete information and deterministic variables. Proactive scheduling is the process of developing a robust baseline schedule by considering the uncertainties. For this purpose, time buffers are inserted into the schedule to protect it from the potential disruptions. In this paper, a method for proactive scheduling of construction projects is proposed based on fuzzy logic. Fuzzy logic is used in the proposed method for representing the uncertainties, because, subjective assessment and lack of data are inherent in many aspects of construction projects. A new approach that uses fuzzy discrete event simulation is proposed for developing the proactive schedule. A numerical example is used to illustrate the developed method, and its performance is evaluated based on the available robust scheduling methods found in the literature.

1. Introduction

According to the literature, many construction operations fail to meet their time and budget (Al-Bahar and Crandall 1990; Assaf and Al-Hejji 2006). Inappropriate planning

[3] Appendix B is published in Proceedings, 3rd International/9th Construction Specialty Conference, CSCE, Ottawa, Ont., June 14-17: CN-007-1-CN-007-11

and scheduling have been identified as the major cause of project delays, due to various uncertain factors that affect construction projects (Assaf and Al-Hejji 2006; Mulholland and Christian 1999). The literature contains both quantitative and qualitative (fuzzy) methods for identifying uncertain factors and estimating these uncertainties (Mulholland and Christian 1999; Ben-David and Raz 2001; Carr and Tah 2001) However, few approaches are proposed for construction for effectively considering the uncertainties within the schedule (Schatteman et al. 2008; Park and Peña-Mora 2004). Recently, proactive scheduling is suggested for construction projects in order to minimize the effect of uncertainties and to provide a robust schedule (Schatteman et al. 2008). In this paper, we propose a method of proactive scheduling that considers fuzzy durations for the activities in the project network. Fuzzy durations are able to factor in uncertainties that result from subjectivity and lack of historical data, both of which are very common in the construction industry.

Proactive scheduling is the process of developing a stable baseline schedule by considering project uncertainties. A stable schedule is one that is protected against project disruptions as much as possible. Proactive scheduling is usually used with a reactive scheduling procedure. The reactive procedure updates the schedule when the project deviates from the baseline schedule (Van de Vonder et al. 2006). Schatteman et al. (2008) have proposed the use of heuristic procedures for providing stable baseline schedules for proactive scheduling of construction projects (Van De Vonder et al. 2006; Van de Vonder et al. 2008). However, the proposed approaches are based on considering probability distributions for the activity durations of the project, and are not able to consider fuzzy activity durations.

On the other hand, when confronted with subjectivity and lack of data, which is common in construction projects, fuzzy set theory is a good alternative for representing the project activity durations. It provides a methodology for handling linguistically expressed and subjective variables. Fuzzy methods have been used successfully in various types of construction projects. For example, fuzzy if-then rules (Zadeh 1973; Mamdani 1974) are used for project scheduling (Ayyub and Haldar 1984), and for predicting industrial construction labour productivity (Fayek and Oduba 2005). Also, fuzzy variables are used to estimate the uncertainty in construction activities (Zhang et al. 2005; Shaheen et al. 2007; Sadeghi et al. 2010).

This paper is organized as follows: 1) an introduction to fuzzy set theory is provided; 2) an algorithm for finding the fuzzy start times of the activities of a project network using fuzzy discrete event simulation is explained; 3) a methodology for developing the stable baseline schedule is discussed (This methodology is based on the obtained fuzzy start times of the activities.); 4) the proposed approach is illustrated with a numerical example, and the results are compared with available proactive scheduling approaches; 5) conclusions and future research are discussed.

2. Fuzzy set theory

A fuzzy set $\tilde{a}$ is defined on the universal set U by assigning a membership degree between 0 and 1 to each member of U. The membership degree indicates the degree to which each element of U is compatible with the properties of the fuzzy set (Zadeh 1965). A fuzzy variable is a fuzzy set that is defined on the real line R. The membership function of a fuzzy variable $\tilde{a}$ is denoted as $\mu_{\tilde{a}}$. For x equal to any possible value of $\tilde{a}$,

$\mu_{\tilde{a}}(x)$ represents the possibility (degree of membership) of x in $\tilde{a}$. An example of a membership function is shown in Figure 1.

The alpha-cut of a fuzzy set $\tilde{a}$ at the level of $\alpha \in (0,1]$ is a set $\tilde{a}_{\alpha}$, whose members have a membership degree equal to or greater than α (Figure 1). The strong alpha-cut of a fuzzy set $\tilde{a}$ at the level of $\alpha \in [0,1]$ is defined as set $\tilde{a}_{\alpha+}$ whose members have a membership degree greater than α. The support of a fuzzy set is equal to its strong alpha-cut at the level of $\alpha = 0$. The core of a fuzzy set is equal to its alpha-cut at the level of $\alpha = 1$.

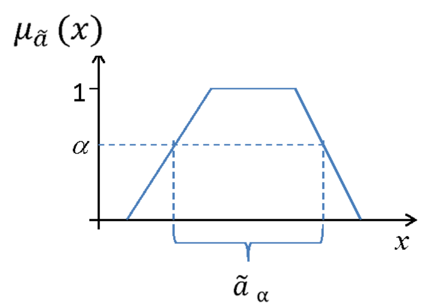


Figure 1 : The alpha-cut of a fuzzy set.

A fuzzy number is a fuzzy variable that has a bounded support and a non-empty core, and its membership function is continuous and convex. A crisp value v is a special case of a fuzzy number: its membership function $\mu_v(x)$ is equal to 0 for $v \neq x$ and is 1 for $v = x$. A triangular fuzzy number is a common fuzzy number used in the literature. It is represented with three values, tri(a,b,c), and its membership function is defined as shown in Equation 1.

$$\mu_{tir(a,b,c)}(x) = \begin{cases} \frac{x-a}{b-a} & a \leq x < b \\ \frac{c-x}{c-b} & b \leq x < c \\ 0 & x < a \text{ or } x \geq c \end{cases} \tag{1}$$

Fuzzy addition, $\oplus$, can be performed on fuzzy numbers using the alpha-cut method. Assume $\tilde{a}$ and $\tilde{b}$ are two fuzzy numbers and $\tilde{c}$ is the fuzzy set of their sum, $\tilde{c} = \tilde{a} \oplus \tilde{b}$, the alpha-cut of $\tilde{c}$ can be calculated at any level of α, using the alpha-cut of $\tilde{a}$ and $\tilde{b}$ (Equation 2).

$$\tilde{c}_{\alpha} = \left(\tilde{a}_{\alpha} + \tilde{b}_{\alpha}\right) \tag{2}$$

Therefore, Equation 3 can be used to construct the fuzzy set $\tilde{c}$ from its alpha-cuts. In this equation the union (U) is performed for the values of $\alpha \in (0,1]$.

$$\tilde{c} = \cup_{\alpha}\left(\tilde{a}_{\alpha} + \tilde{b}_{\alpha}\right).\alpha \tag{3}$$

A comparison of two fuzzy numbers may be required for decision making purposes. Various fuzzy ranking methods are suggested for comparing fuzzy numbers (Zhang et al. 2005; Bortolan and Degani 1985; Chen 1985; Chen et al. 1992; Liou and Wang 1992; Tran and Duckstein 2002; Perrone et al. 2001). The simplest method is to defuzzify the fuzzy numbers and compare their defuzzified values. Defuzzification is used to convert a fuzzy variable into a crisp value. The centroid method is one common method for defuzzification. In this method, the defuzzified value is calculated by finding the center of mass of the membership function.

3. Fuzzy Discrete Event Simulation (FDES)

In this section, an algorithm is proposed for generating the start and end times for a resource-constrained project network with fuzzy activity durations. This problem is referred to as FRCPN (fuzzy resource constrained project network). Wang (2004) proposes an algorithm for finding the fuzzy start and end times for FRCPN. Other methods such as fuzzy PERT and fuzzy Monte Carlo simulation can be used for finding the fuzzy start and end times of activities. In this paper, fuzzy discrete event simulation (FDES) is proposed to solve this problem. The proposed FDES method benefits from the latest advancements in FDES.

A FDES is a discrete event simulation in which time variables in the simulation (e.g., inter-arrival times, times between failures, durations of the activities) are fuzzy variables instead of probability distributions (Perrone et al. 2001). Recently, FDES has been proposed for considering subjective and linguistically expressed data in the simulation of construction projects (Zhang et al. 2005).

A project can be modeled using an activity-on-node network, G = (V, E). The nodes E represent the activities (0,…n), where 0 is the start dummy activity and n is the end dummy activity; the arcs V represent precedence relations. The set of available resources are defined as R, and set of required resources are denoted by Q, where the required resources of activity i for resource j is denoted by Qij. Also, the duration of activity j is a fuzzy number $\widetilde{d}_j$.

FDES is performed to find the fuzzy start times of the activities for the FRCPN. The simulation time $\widetilde{T-now}$ is in the form of a fuzzy number in discrete event simulation. At the start of the simulation, $\widetilde{T-now}$ is equal to 0, and the event list is empty. At each point in time, $\widetilde{T-now}$ activities that can be started are identified, based on the availability of the resources and their predecessors. An event is created for each of these activities, and its required resources are captured. Therefore, the event list in the FDES contains the activities of the network that are eligible to start. Each event in the event list corresponds to an activity in the project network. Here, the policy is to start each activity as soon as possible. When activities are competing for resources, the priorities of the activities are used to decide which activity should be fired first. These priorities are assigned to the activities before the simulation.

To create an event for activity j at time $T \widetilde{- now}$, the event time $\tilde{t}_j$ is calculated by summing the activity duration $\tilde{d}_j$ and $T \widetilde{- now}$. Since both of these values are in the form of fuzzy numbers, fuzzy addition must be performed to find the event time. Therefore, the start time of the activity j is recorded as $T \widetilde{- now}$, and the finish time of the activity is $\tilde{t}_j$ (Equation 4). As a result of the fuzzy addition, each event includes the event time that represents the completion time of its activity.

$$\tilde{t}_j = \tilde{d}_j \oplus T \widetilde{- now} \tag{4}$$

After generating events for all eligible activities, $T \widetilde{- now}$ is updated based on the minimum (min) event time in the event list. Given k events in the event list with event times $\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_k$, $T \widetilde{- now}$ is calculated using Equation 5.

$$T \widetilde{- now} = min(\tilde{t}_1, \tilde{t}_1, \dots, \tilde{t}_1) \tag{5}$$

Since the event times $\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_k$ are fuzzy sets, fuzzy ranking must be performed to find the smallest event. Various fuzzy ranking methods have been used for fuzzy discrete event simulation in the literature (Zhang et al. 2005; Perrone et al. 2001). However, no single ranking method has been proven to be the best for fuzzy discrete event simulation. In this research, the defuzzification-based ranking method is used, which is one of the simplest methods that exist in the literature (Perrone et al. 2001). The centroid method is first used to defuzzify the fuzzy sets, after which they are ranked based on their defuzzified values. For future research, different ranking methods will be explored for the proposed approach.

The event with the smallest time is removed from the event list, and the resources captured by the activity of that event are released. Again, the activities that are not yet started are checked for updating the event list. This process continues until all the activities in the project network are finished. At the end of the simulation, $\widetilde{T-now}$ represents the possibility distribution of the project completion time.

Using this procedure, a schedule for FRCPN can be developed that indicates the fuzzy start times, $\widetilde{s'}_0, \widetilde{s'}_1, \dots, \widetilde{s'}_n$, for each activity. These start times are based on starting each activity as soon as possible. Therefore, s'_0 is always equal to 0. Also, s'_n represents the fuzzy set of the project completion time, because activity n is the end dummy activity and its duration is 0. The resulting schedule is used as the basis for developing the stable baseline schedule for FRCPN.

4. Developing the Robust Baseline Schedule

In this section, an algorithm is proposed for developing a robust schedule for FRCPN. The goal is to develop a baseline schedule S for FRCPN that is protected against disruptions as much as possible. In proactive scheduling, the start times of the activities $s_0, s_1, \dots, s_n$ are defined as crisp values to provide a robust schedule. This robustness is calculated based on the comparison between the real start time of the activity during the execution of the project and its scheduled start time. In order to maintain stability of the project, each activity is only allowed to start after its scheduled start time (Van de Vonder et al. 2006), which is in contrast with the policy of starting each activity as soon as possible. This approach is especially practical for construction projects, in which the activities are subcontracted or materials are ordered in advance in order to be available on site on the scheduled start times.

First, a robustness measure is defined for an activity when we have a fuzzy type of uncertainty. In the proposed approach, the pessimistic criterion of Dubois and Prade (1999) is used for finding the robustness measure. Wang (2004) used the same criterion when both the preferred start (finish) time of an activity j, $\widetilde{s_j}$, and its calculated start (finish) time, $\widetilde{s_j}$, are fuzzy sets. Equation 6 indicates the robustness measure of activity j, RM_j, that is used by Wang (2004), where inf stands for infimum, sup stands for supremum, max stands for maximum, and min stands for minimum.

$$RM_j = inf_x \max(1 - \mu_{\widetilde{s_j}}(x), \mu_{\widetilde{s_j}}(x)) \tag{6}$$

In contrast with Wang (2004), in the proposed approach, the preferred start time for the activity j is a crisp number equal to its scheduled start time, s_j, in the baseline schedule. Since crisp values are a special case of a fuzzy set, the value s_j can be defined as a fuzzy set, $\widetilde{s_j}$, with a membership function shown in Equation 7.

$$\mu_{\widetilde{s_j}}(x) = \begin{cases} 1 & x = s_j \\ 0 & otherwise \end{cases} \tag{7}$$

According to Equation 7, the value of $\max(1-\mu_{\widetilde{S_j}}(x), \mu_{\widetilde{S_j}}(x))$ is equal to 1 for $x = s_j$. Also, $\max(1-\mu_{\widetilde{S_j}}(x), \mu_{\widetilde{S_j}}(x))$ is equal to $1-\mu_{\widetilde{S_j}}(x)$ for $x \neq s_j$. Therefore, Equation 6 can be rewritten as shown in Equation 8.

$$RM_j = \min(inf_{x\neq s_j} \max\left(1-\mu_{\widetilde{S_j}}(x), 0\right), \max\left(1-\mu_{\widetilde{S_j}}(s_j), 1\right)) = 1 - sup_{x\neq s_j}\mu_{\widetilde{S_j}}(x) \qquad (8)$$

Assume the start time of activity j, based on the availability of resources and predecessors, is obtained as $\widetilde{s''_J}$. Fuzzy addition is performed to find the membership function of $\widetilde{s''_J}$, $\mu_{\widetilde{s''_J}}(x)$, within the FDES framework. However, the fuzzy start time of the activity in the proactive scheduling approach, $\widetilde{S_J}$, is not equal to $\widetilde{s''_J}$. Based on the scheduled start time, s_j, of activity j, activity j can only start after the allowed time of s_j.Therefore, the possibility that activity j starts at time x, $\mu_{\widetilde{S_j}}(x)$, for $x < s_j$ is 0. Activity j will start at the scheduled time s_j, with the possibility that all the predecessors are ready before time s_j. Therefore, $\mu_{\widetilde{S_j}}(s_j) = Poss\left(s_j > \widetilde{s''_J}\right)$. Also, the possibility of the start of the activity at time $x > s_j$ is equal to the possibility of having all the predecessors and resources ready, $\mu_{\widetilde{s''_J}}(x)$. The membership function for the start time of an activity in a proactive schedule is shown in Equation 9.

$$\mu_{\widetilde{S_j}} = \begin{cases} Poss\left(s_j > \widetilde{s''_J}\right) & x = s_j \\ \mu_{\widetilde{s''_J}}(x) & x > s_j \\ 0 & x < s_j \end{cases} \qquad (9)$$

Therefore, Equation 8 can be rewritten as shown in Equation 10. Also, According to Equation 9, the upper limit of the alpha-cut of the start time of the activity at $\alpha = \mu_{\widetilde{s_j}}(x)$ for $x > s_j$ is equal to the upper limit of the alpha-cut of $\widetilde{s''_j}$ at the same alpha level.

$$RM_j = \min(inf_{x>s_j} \max\left(1 - \mu_{\tilde{s}_j}(x), 0\right), \max\left(1 - \mu_{\tilde{s}_j}(s_j), 1\right)) = 1 - sup_{x>s_j}\mu_{\widetilde{s_j}}(x) \quad (10)$$

To find the robust baseline schedule, the fuzzy start times obtained based on the algorithm of Section 3 are used. Assume $\widetilde{s'_0}, \widetilde{s'_1}, \ldots, \widetilde{s'_n}$ are the start times of the activities (0,...,*n*) obtained by starting each activity as soon as possible using FDES. Starting from the end dummy activity n, the value $\widetilde{s'_n}$ is equal to the finish time of the project. Therefore, the start time of activity n, s_n in the proactive schedule is equal to or less than the project due date δ_n: $s_n \le \delta_n$. This is because we cannot schedule the project to finish after the project deadline (due date). Also, the start time of activity n in the proactive scheduling model, $\widetilde{s_n}$, is greater than or equal to its start time based on starting each activity as soon as possible, $\widetilde{s'_n}$. Therefore, Equation 11 represents the robustness measure of activity n, RM_n, based on $\widetilde{s'_n}$.

$$RM_n \le 1 - sup_{x>\delta_n}\mu_{\widetilde{s'_n}}(x) \quad (11)$$

Assume α is equal to $\mu_{\widetilde{s'_n}}(\delta_n)$. Five conditions may occur when comparing the due date δ_n and alpha-cut of $\widetilde{s'_n}$ at the level of α (Figure 2): a) $\alpha = 0$ and is less than the support of the $\widetilde{s'_n}$; b) $\alpha = lowerlimit\left(\widetilde{s'_n}\right)_\alpha$; c) $\alpha = 1$ (d) $\alpha = upperlimit\left(\widetilde{s'_n}\right)_\alpha$; e) $\alpha = 0$ and is greater than the support of $\widetilde{s'_n}$. For conditions (a), (b), and (c), according to Equation 11, the robustness measure of the last activity is 0. In these cases, the project due date is not realistic for this project, and a robust schedule is not possible.

For the conditions (d) and (e), the robustness measure of the last activity is less than or equal to $1 - \alpha$ (see Equation 11). In this case, the start time, s_j, of activity j, in the baseline schedule is calculated as the upper limit of the strong alpha-cut of $\widetilde{s'_J}$ at the level of α (Equation 12).

$$s_j = upperlimit(\widetilde{s'_J})_{\alpha+} \qquad (12)$$

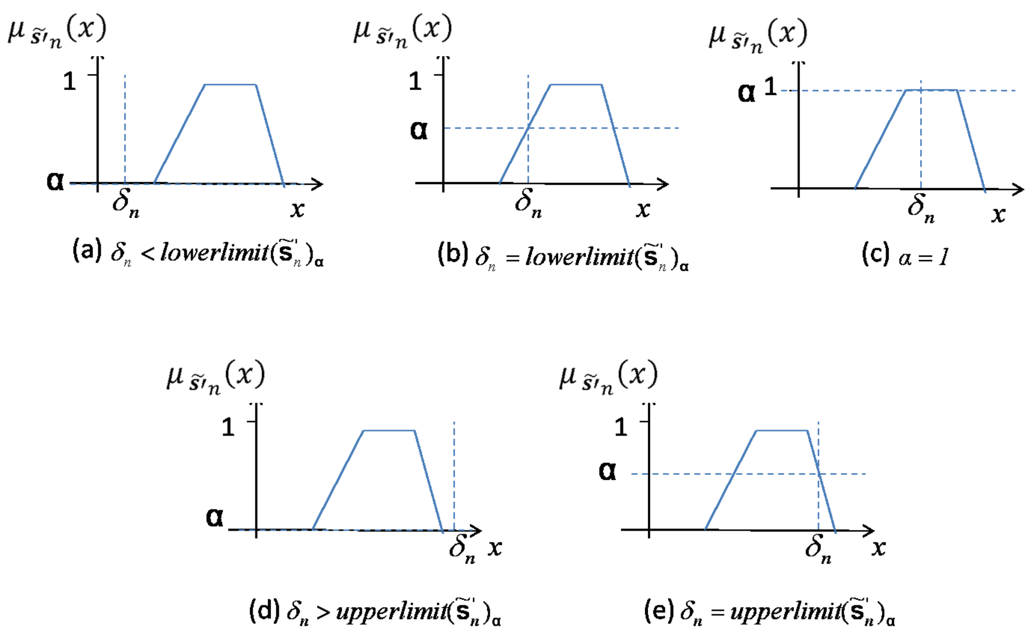


Figure 1: Different conditions between the membership functions of the project completion time based on the policy of starting each activity as soon as possible, $\boldsymbol{\mu_{\widetilde{s'}_n}(x)}$, and the project due date, $\boldsymbol{\delta_n}$.

By calculating s_j, the real start time of activity j, $\widetilde{s_J}$ can be calculated based on Equation 9. Equation 9 maintains the upper limit of the alpha-cut at the level of $\alpha = \mu_{\widetilde{S_J}}(x)$, for

$x > s_j$. Also, the upper limit of the alpha-cut of the result of the addition of two fuzzy sets, $\tilde{a}$ and $\tilde{b}$, only depends on the upper limit of the alpha-cuts of $\tilde{a}$ and $\tilde{b}$ (see Equation 2). Therefore, the upper limit of the alpha-cut of the start time of the activities in the proactive schedule for values of $\alpha = \mu_{\widetilde{s_j}}(x)$ for $x > s_j$ does not change compared to the case of starting each activity as soon as possible: $upperlimit(\widetilde{s_j}\ )_\alpha = upperlimit(\widetilde{s'_j})_\alpha$. As a result, $\mu_{\widetilde{s_j}}(x) = \mu_{\widetilde{s'_j}}(x)$ for $x > s_j$. Therefore the robustness measure of activity j, RM_j, can be calculated as $1 - sup_{x>s_j}\mu_{\widetilde{s'_j}}(x)$. According to Equation 12, $1 - sup_{x>s_j}\mu_{\widetilde{s'_j}}(x) = 1 - \alpha$. Since α is equal to $\mu_{\widetilde{s'_n}}(\delta_n)$ and does not depend on the activity index j, the robustness measures of all of the activities in the project network are equal. The proposed approach maximizes the robustness measure of the project completion time, because, according to Equation 11, $RM_n \leq 1 - \alpha$, and the proposed approach results in $RM_n = 1 - \alpha$.

5. Numerical Example

The example in Van de Vonder et al. (2006) is used in this section to illustrate the proposed proactive scheduling approach. A project network with eight non-dummy activities is considered. One type of resource is defined for the project with an availability of 10 units. The due date of the project is at time 20. For the durations of the activities, Van de Vonder et al. (2006) provided the minimum (min), maximum (max), and mean for each activity. Then, they defined a beta distribution for the duration of each activity based on the provided values for min, max and mean. Here, triangular fuzzy numbers are used, tri(min,mean,max), for the durations of the activities using the min, max, and mean values defined by Van de Vonder et al. (2006) for each activity. The activity on node (AON) approach is used to represent the project network (Figure 3). Table 1 indicates the duration of each activity and its required resources. Activities 0 and 9 are start and end dummy activities, respectively, in the project network.

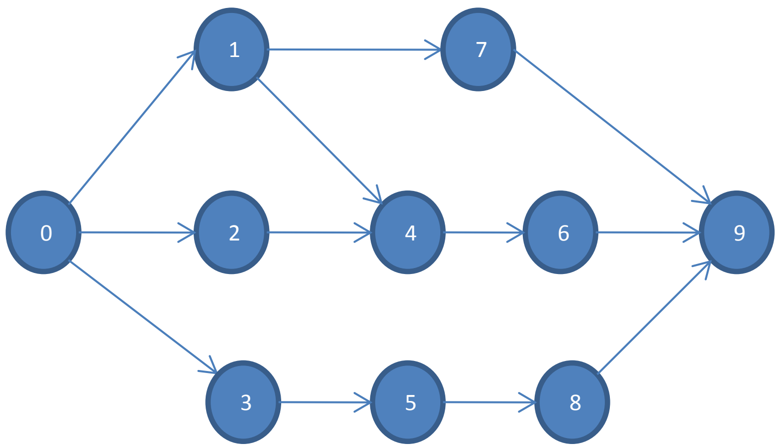


Figure 2: The project network for the numerical example.

Table 1 The duration of each activity and its required resources.

Activity number(j)	Activity duration($\tilde{d}_j$)	Required resources(r_j)
0	tri(0,0,0)	0
1	tri(2,4,9)	5
2	tri(3.75, 5,8.125)	3
3	tri(1,2,4.5)	4
4	tri(1,4,11.5)	4
5	tri(2.5,5,11.25)	3
6	tri(3,4,6.5)	5
7	tri(0.5,2,5.75)	3
8	tri(1.5,2,3.25)	6
9	tri(0,0,0)	0

The FDES approach is used to find the fuzzy sets of the start and finish times of the activities, $\widetilde{s'}_1, \widetilde{s'}_2, \ldots, \widetilde{s'}_9$. Table 2 indicates these fuzzy start times. For the next step, α is calculated as the membership degree of the due date in the last dummy activity: α =

$\mu_{\widetilde{s'_9}}(20)$. The membership function of $\widetilde{s'_9}$ is equal to tri(9.25,15,29.37) (Table 2). Therefore, according to Equation 1, the membership degree of 20 is equal to (29.37-20)/(29.37-15)=0.65. Finally the start time of each activity j is found as $\mathrm{upperlimit}(\widetilde{s'_j})_{0.65+}$. Table 2, indicates the start times obtained for each activity.

Table 2 Fuzzy start times of the activities $\widetilde{s'_1}, \widetilde{s'_2}, \dots, \widetilde{s'_9}$ obtained from the FDES and their scheduled start times $s_1, s_2, \dots, s_9$ for the baseline schedule.

Act(j)	Fuzzy start time($\widetilde{s'_j}$)	Scheduled start time(s_j)
1	tri(0,0,0)	0
2	tri(0,0,0)	0
3	tri(2,4,9)	5.7
4	tri(3.75,5,8.12)	6.0
5	tri(3,6,13.5)	8.6
6	tri(4.75,9,19.62)	12.6
7	tri(3,6,13.5)	8.6
8	tri(7.75,13,26.12)	17.5
9	tri(9.25,15,29.37)	20.0

The results of the proposed approach are compared with available heuristic methods that exist in the literature for robust scheduling of a probabilistic problem. The robust baseline schedules were developed by Van De Vonder et al. (2006) for the probabilistic version of this example. Heuristic methods that are used for developing these schedules are RFDFF (resource flow dependant float factor), VADE (virtual activity duration extension heuristic), and STC (starting time criticality). Table 3 indicates the start times of the activities in the proactive schedule using each of the methods (Van de Vonder et al. 2006). Table 3 also indicates the start times in the traditional scheduling method, in which the uncertainties are not considered and the most likely values of the activity durations are used for scheduling. The start time of activity j is represented by

$s_j^{RFDFF}, s_j^{VADE}, s_j^{STC}, s_j^{Tradional}$ for the RFDFF, VADE, STC and traditional scheduling methods, respectively.

Table 3 The start time of the activities in the example using different heuristic methods (Van de Vonder et al. 2006).

Act(j)	s_j^{RFDFF}	s_j^{VADE}	s_j^{STC}	$s_j^{Tradional}$
1	0	0	0	0
2	0	0	0	0
3	4	6	5	4
4	5	6	6	5
5	8	9	8	6
6	10	13	11	9
7	7	9	8	6
8	15	18	17	13
9	20	20	20	20

To compare the schedule resulting from the proposed proactive scheduling method, S, and the schedule resulting from other method, S^{RFDFF}, S^{VADE}, S^{STC},and $S^{Traditional}$, the distance (Δ) between each of the two schedules is calculated. This distance is calculated as the sum of the absolute (abs) difference between the scheduled start time of the activities in the proposed method, and the schedule start time by the other method. For example, the distance between the proposed schedule and the STC approach, $\Delta(S, S^{STC})$ is calculated as $\sum_{j=0}^{n} abs(s_j^{STC} - s_j)$. Table 4 indicates the distance between the proposed schedule and each of the schedules resulting from other approaches (RFDFF, VADE, STC, traditional scheduling).

Table 4 The distance between the proposed proactive schedule S and schedules S^{RFDFF}, S^{VADE}, S^{STC}, and $S^{Traditional}$.

$\Delta(\boldsymbol{S}, \boldsymbol{S}^{\boldsymbol{RFDFF}})$	$\Delta(\boldsymbol{S}, \boldsymbol{S}^{\boldsymbol{VADE}})$	$\Delta(\boldsymbol{S}, \boldsymbol{S}^{\boldsymbol{STC}})$	$\Delta(\boldsymbol{S}, \boldsymbol{S}^{\boldsymbol{Traditional}})$
10.27	1.90	4.27	16.27

As indicated in Table 4, the results from the proposed scheduling approach are closer to the results from the proactive scheduling approaches based on probabilistic durations than the results from the traditional scheduling approach. In fact, the start time, s_j, of each activity j in the proposed schedule is always between the start time of the activity in the VADE and STC methods: $s_j^{STC} < s_j < s_j^{VADE}$ (see Tables 2 and 3). Therefore, the schedules resulting from the proposed approach are between the schedules resulting from the STC and VADE approaches. VADE and STC have both been proven to be appropriate proactive heuristics that have higher stability compared to the traditional scheduling method (Van de Vonder et al. 2008). Therefore, the results of the fuzzy robust scheduling method are comparable to the results of proactive scheduling approaches developed for project networks with probabilistic activity durations. The main advantage of the proposed approach is in its ability to develop a proactive schedule when dealing with subjectivity and linguistically expressed information (both of which are common in construction), by using fuzzy numbers to represent activity durations rather than probabilistic distributions.

6. Conclusions and Future Research

In this paper, a new approach for developing a stable project baseline schedule for a FRCPN (fuzzy resource constrained project network) is proposed. For this purpose, a fuzzy discrete event simulation algorithm is developed to solve the fuzzy resource constrained project network and to find the fuzzy start and finish times for each activity. Then, alpha-cuts of the fuzzy start times of the activities are used to develop the stable schedule. Based on the defined robustness measure in this paper, the proposed schedule maximizes the robustness of the project completion time. It also provides equal robustness for all of the activities in the project network. The results of the numerical example indicate that the schedule resulting from the proposed method is between the schedules resulting from two reliable proactive scheduling approaches that incorporate probabilistic uncertainty, namely the STC and VADE approaches (Van de Vonder et al. 2006). The proposed proactive scheduling approach can be very useful for providing a robust baseline schedule for construction projects, especially when fuzzy numbers are used for representing the durations of the activities (due to the lack of data and subjectivity). This research can be developed further in the following areas:

- A weight may be considered for each activity to define the importance of the robustness of each activity compared to the other activities. This weight depends on the penalties and costs that occur if the project deviates from the scheduled start time of an activity. For example, the robustness of an activity that is subcontracted is more important than the activities that are performed by the organization itself.

- The FDES approach that is used for finding the fuzzy start times of the activities can be improved. For example, different ranking methods should be explored for FDES.

- In future research, proactive schedules can be developed using fuzzy PERT or fuzzy Monte Carlo simulation instead of FDES. The results of these approaches can be compared with the results of the proposed methodology.

- Case studies and/or computer experiments should be performed to investigate the effects of the proposed approach on decreasing the costs and delays in construction projects.

7. References

Al-Bahar, J. F., & Crandall, K. C. (1990). Systematic risk management approach for construction projects. Journal of Construction Engineering and Management, 116(3), 533-546.

Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. International journal of project management, 24(4), 349-357.

Ayyub, B. M., & Haldar, A. (1984). Project scheduling using fuzzy set concepts. Journal of Construction Engineering and Management, 110(2), 189-204.

Ben-David, I., & Raz, T. (2001). An integrated approach for risk response development in project planning. Journal of the Operational Research Society, 14-25.

Bortolan, G., & Degani, R. (1985). A review of some methods for ranking fuzzy subsets. Fuzzy sets and Systems, 15(1), 1-19.

Carr, V., & Tah, J. H. M. (2001). A fuzzy approach to construction project risk assessment and analysis: construction project risk management system. Advances in Engineering software, 32(10), 847-857.

Chen, S. H. (1985). Ranking fuzzy numbers with maximizing set and minimizing set. Fuzzy sets and Systems, 17(2), 113-129.

Chen, S. J. J., Hwang, C. L., Beckmann, M. J., & Krelle, W. (1992). Fuzzy multiple attribute decision making: methods and applications. Springer-Verlag New York, Inc.

Dubois, D., & Prade, H. (1999). Qualitative possibility theory and its applications to constraint satisfaction and decision under uncertainty. International Journal of Intelligent Systems, 14(1), 45-61.

Fayek, A. R., & Oduba, A. (2005). Predicting industrial construction labor productivity using fuzzy expert systems. Journal of construction engineering and management, 131(8), 938-941.

Liou, T. S., & Wang, M. J. J. (1992). Ranking fuzzy numbers with integral value. Fuzzy sets and systems, 50(3), 247-255.

Liu, B. (2009). Theory and practice of uncertain programming. Vol. 239, Springer.

Mamdani, E. H. (1974). Application of fuzzy algorithms for control of simple dynamic plant. Proceedings of the Institution of Electrical Engineers, 121(12), pp. 1585-1588, IET Digital Library.

Mulholland, B., & Christian, J. (1999). Risk assessment in construction schedules. Journal of Construction Engineering and Management, 125(1), 8-15.

Park, M., & Peña-Mora, F. (2004). Reliability buffering for construction projects. Journal of construction engineering and management, 130(5), 626-637.

Perrone, G., Zinno, A., & La Diega, N. (2001). Fuzzy discrete event simulation: A new tool for rapid analysis of production systems under vague information. Journal of Intelligent Manufacturing, 12(3), 309-326.

Sadeghi, N., Fayek, A. R., & Pedrycz, W. (2010). Fuzzy Monte Carlo simulation and risk assessment in construction. Computer-Aided Civil and Infrastructure Engineering, 25(4), 238-252.

Schatteman, D., Herroelen, W., Van de Vonder, S., & Boone, A. (2008). Methodology for integrated risk management and proactive scheduling of construction projects. Journal of Construction Engineering and Management, 134(11), 885-893.

Shaheen, A. A., Fayek, A. R., & AbouRizk, S. M. (2007). Fuzzy numbers in cost range estimating. Journal of Construction Engineering and Management, 133(4), 325-334.

Tran, L., and Duckstein, L. (2002). Comparison of fuzzy numbers using a fuzzy distance measure. Fuzzy Sets and Syst., 130(3), 331–341.

Van de Vonder, S., Demeulemeester, E., Leus, R., & Herroelen, W. (2006). Proactive-reactive project scheduling trade-offs and procedures. Perspectives in Modern Project Scheduling, pp. 25-51, Springer US.

Van de Vonder, S., Demeulemeester, E., & Herroelen, W. (2008). Proactive heuristic procedures for robust project scheduling: An experimental analysis. European Journal of Operational Research, 189(3), 723-733.

Wang, J. (2004). A fuzzy robust scheduling approach for product development projects. European Journal of Operational Research, 152(1), 180-194.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics, (1), 28-44.

Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.

Zhang, H., Tam, C. M., & Li, H. (2005). Modeling uncertain activity duration by fuzzy number and discrete-event simulation. European journal of operational research, 164(3), 715-729.