Development of Fuzzy Hybrid Approaches to Project Delivery Method Selection in Highway Construction

By

Phuong H.D. Nguyen
M.S., California State University, East Bay, 2016
B.S., Vietnam National University, HCM University of Technology, 2014

Submitted to the graduate degree program in the Department of Civil, Environmental, and Architectural Engineering and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Co-Chair: Dr. Dan Tran

____
Co-Chair: Dr. Brian Lines

____
Dr. Joshua Roundy

____
Dr. Elaina Sutley

____
Dr. Guanghui (Richard) Wang

____
Dr. Long D. Nguyen

Date Defended: July 2nd, 2020

The dissertation committee for Phuong Nguyen certifies that this is the approved version of the following dissertation:

Development of Fuzzy Hybrid Approaches to Project Delivery Method Selection in Highway Construction

Co-Chair: Dr. Dan Tran

Co-Chair: Dr. Brian Lines

Date Approved: July 14th, 2020

ABSTRACT

Selection of project delivery methods is a success factor in delivering highway construction projects because it has a substantial impact on the project performance, such as cost, time, and quality. Project delivery decision-making processes have been heavily relied on experts' opinions and subjective judgements of professionals to evaluate quantitative and qualitative decision variables. Although current quantitative and probabilistic methods provide a robust means to analyze quantitative variables, they are not ideally suited for treating uncertainties encountered in qualitative variables. Fuzzy set theory is a mathematical approach that can accommodate a combination of quantitative and qualitative variables. This dissertation aimed at investigating the applications of fuzzy set theory and fuzzy logic to support decision-making processes in project delivery method selections. Using an empirical dataset of 254 completed highway construction projects, three fuzzy-based applications, including fuzzy cluster analysis, fuzzy pattern recognition, and fuzzy Bayesian inference system were developed, trained, and tested. As a result, fuzzy cluster analysis was used to establish seven common project clusters that share high similarities in project characteristics, project complexity, delivery risks, cost growth, and project delivery methods. Fuzzy pattern recognition was used to develop a fuzzy rule-based inference system based on the seven identified project clusters to help recognize an appropriate project delivery method associated with potential cost growth for new highway projects. Fuzzy Bayesian networks were used to develop the theoretical framework of fuzzy Bayesian inference system which is able to depict the causal relationships between project characteristics, project complexity, delivery risks, and project delivery methods. The flexibility of fuzzy membership functions in the developed applications helps leverage the evaluation of a combination of quantitative and qualitative variables in highway project delivery method

selection. In addition, these data-driven fuzzy applications also allow for multiple decision scenarios based on the decision maker's judgements of delivery risks and project complexity. This dissertation contributes to the body of knowledge by demonstrating quantitative approaches derived from fuzzy set theory and fuzzy logic to support the selection of project delivery methods in highway construction. Additionally, the results from the developed fuzzy-based applications also provide insights regarding cost performance comparisons between project delivery methods. This study may assist highway agencies in making project delivery decisions based on project attributes, historical data and their relevant experience.

Dissertation Format

This dissertation follows the three-journal-paper format. First, Chapter 1 includes the research problems and objectives shown in the Introduction section followed by Research Methodology with details regarding each research phase. Second, Chapters 2, 3, and 4 are formed in terms of three journal papers. Specifically, Chapter 2 demonstrates the establishment of seven empirical clusters of highway projects that share commonalities in project attributes, such as facility type, project type, project complexity, delivery risks, cost performance, and delivery methods used, based on fuzzy cluster analysis; this chapter provided the first peer-reviewed journal article, published in the American Society of Civil Engineers (ASCE) *Journal of Construction Engineering and Management (JCEM)*. Chapter 3 presents an empirical rule-based inference system based on fuzzy pattern recognition to identify project delivery methods for new highway projects; this chapter produced the second journal article, which has been submitted and under review with ASCE *JCEM*. Chapter 4 illustrates the development of a theoretical framework of fuzzy Bayesian rule-based inference system, which is expected to leverage the established fuzzy inference system to be a decision-aid tool for highway agencies to selecting appropriate delivery methods; this chapter will produce the third journal paper which will be submitted to ASCE *JCEM*. Finally, the Conclusion section summarizes the entire dissertation and restates the contributions to the body of knowledge. Appendices, where all related tables, figures, graphs, algorithms, and pseudo R programming codes used in this dissertation are also combined at the end. Specifically, Appendix A includes the exploratory factor analysis of 31 delivery risks, determination of project clusters, fuzzy C-means cluster analysis algorithm, and other information of fuzzy cluster analysis. Appendix B includes fuzzy membership functions (Gaussian type), rule-based inference formulation, and programming graphical user interface.

To my family and friends

ACKNOWLEDGMENTS

I started my doctoral study journey by joining the Department of Civil, Environmental, and Architectural Engineering (CEAE) at the University of Kansas (KU) in the Fall 2016. I really appreciate the helps from my supervisors, my dissertation committee members, CEAE faculty and staffs, my family, and my friends. At the end of the day, I am very proud to be a Jayhawk.

First, I want to express my most sincerely gratitude and respect to my supervisors, Drs. Dan Tran and Brian Lines, who always help me with an incredible vision and inspiration. I am very glad that they continually support me to achieve my research and teaching career goals. They inspire and encourage me to develop my research career in construction engineering and management with publications, research projects, and presentations at well-known construction research conferences. They also provided me with extraordinary opportunities to improve my teaching ability through teaching assistantships in construction-related courses as well as guest lectures at KU.

Second, I am thankful to all my dissertation committee members who provided me insightful comments. I am privileged to get invaluable suggestions and guidance from Drs. Sutley and Roundy from CEAE as well as Dr. Wang from the Department of Electrical Engineering and Computer Science at KU. I would really like to thank Dr. Nguyen from Florida Gulf Coast University for serving as an outside member on my committee. Furthermore, I really appreciate the helps from CEAE Department's faculty and staffs, especially Susan Scott and Reta Sowa for their wonderful help during my time at KU.

Finally, many thanks to my family and friends for their unconditional love and incredible support. I cannot pursue my doctoral degree without their encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 4

CHAPTER 1:

INTRODUCTION

Background

A project delivery method is a framework that affirms contractual relationships between project stakeholders and determines how the project will be executed. Selection of project delivery methods is a success factor of highway projects because it has substantial impacts on project performance, such as cost, time, and quality (Allen 2001). The rise of alternative project delivery methods (APDMs) or alternative contracting methods (ACMs) contributes to optimization of project performance, especially in highway projects, where large amounts of budgets are spent and require strict guarantees (El Asmar et al. 2013; Ghavamifar 2009). On the demand of enhancing productivity of infrastructure construction, such as highways, bridges, roads, and other horizontal projects, many changes and adoptions have been recognized in the project administration and management with project delivery methods (Alleman et al. 2016; Shrestha et al. 2012).

Selecting a project delivery method, one of the critical success factors, directly affects project performance and any other subsequent decisions throughout the project lifecycle. Implementation of different project delivery methods lead to different scenarios of project performance and effectiveness related to project characteristics, identified risks, and expected project accomplishment. Accordingly, there is no "one size fits all" approach to selecting a project delivery method. Rather, each project may have an appropriate delivery method depending on its characteristics and specifications. Suitability of typical project delivery methods in a highway project is exclusively determined based on its unique project characteristics and related subjective ratings of delivery risks.

The current construction literature has recognized that selection of appropriate project delivery methods is essentially based upon experts' experience-based judgements and inspired the use of quantitative approaches to establish project-delivery-selection supportive frameworks and models. Because of uniqueness of construction projects, where each project has its own characteristics, level of project complexity, and associated delivery risks, quantifying project inputs is challenging and requires more empirically-grounded scientific approaches. Although many studies have offered a process and guidance to select the most suitable delivery method, there is a lack of understanding how to rigorously address qualitative criteria. The main gap is that project-delivery-selection research is lacking in identifying a method to provide typical patterns in terms of delivery methods and project performance. Another research gap consists of constraints of qualitative inputs because of subjective judgements. For example, it is difficult to incorporate the qualitative risk inputs and other quantitative variables of project attributes, such as facility types, project types, and complexity, in construction (Creedy 2006; Creedy et al. 2010). The last research gap is the limitation of applications of fuzzy set theory in selection of project delivery methods in terms of assessing project performance, including cost, schedule, and quality, based on typical project characteristics and risk profiles.

Research Problem

Making decisions in the highway sector is a challenging task because of many factors involved. Those factors are represented in terms of different combinations of quantitative and qualitative criteria in construction decision-making processes (Gransberg and Shane 2010; Touran et al. 2011). A range of decisions in the construction industry commonly rely upon qualitative inputs

(Lam et al. 2001); for example, selecting a project delivery method often considers project characteristics, such as risk level, project type, combination of scope categories, level of complexity, and more (Khanzadi et al. 2016). Traditional probabilistic approaches find difficulties in investigating the qualitative criteria (Al Nahyan et al. 2018). Those approaches are unable to either model linguistic expressions or evaluate imprecise concepts and vagueness in input data of decision-making processes (Elbarkouky et al. 2016). In addition to proposing statistical probabilistic models to quantify different sets of decision criteria (Tran and Molenaar 2015), decision-making processes also require systematic approaches to provide empirical comparisons between decision alternatives. The construction industry involves a large number of historical data that contributes necessary evidences to different decision-making scenarios. Historical data, which includes a combination of quantitative and qualitative information, can be used to determine common groups (i.e., clusters) of construction projects based on relevant attributes. Fuzzy set theory is a great means to investigate and model qualitative data using membership functions (Elwood 2014). Other fuzzy hybrid approaches can utilize the clustering results from historic data to aid the selection project delivery methods.

Research Objectives

To address the underlying gaps in the construction literature, this dissertation aims to:

(1) Identify the common groups of highway projects that exhibit high similarities in project characteristics, project complexity, delivery risks, project delivery methods, and cost growth.

(2) Recognize typical patterns of project delivery methods associated with cost growth for new highway projects based on a combination of quantitative and qualitative variables and develop a programming-based graphical user interface to support pattern recognitions in practice.

(3) Investigate the causal relationships between facility type, project type, project size and duration, project complexity, delivery risks, and cost growth in project delivery method selections.

Research Methodology

This dissertation utilized a research framework from the Stanford Center for Integrated Facility Engineering (CIFE). This framework provides scientific researchers a comprehensive technical procedure where conceptual milestones are generated in each step (Lampe 2015). Figure 1 shows a nine-step research framework, including observed problems, intuitions, theoretical points of departure, research methods, research questions, research tasks, validation results, claimed contributions, and predicted impacts.

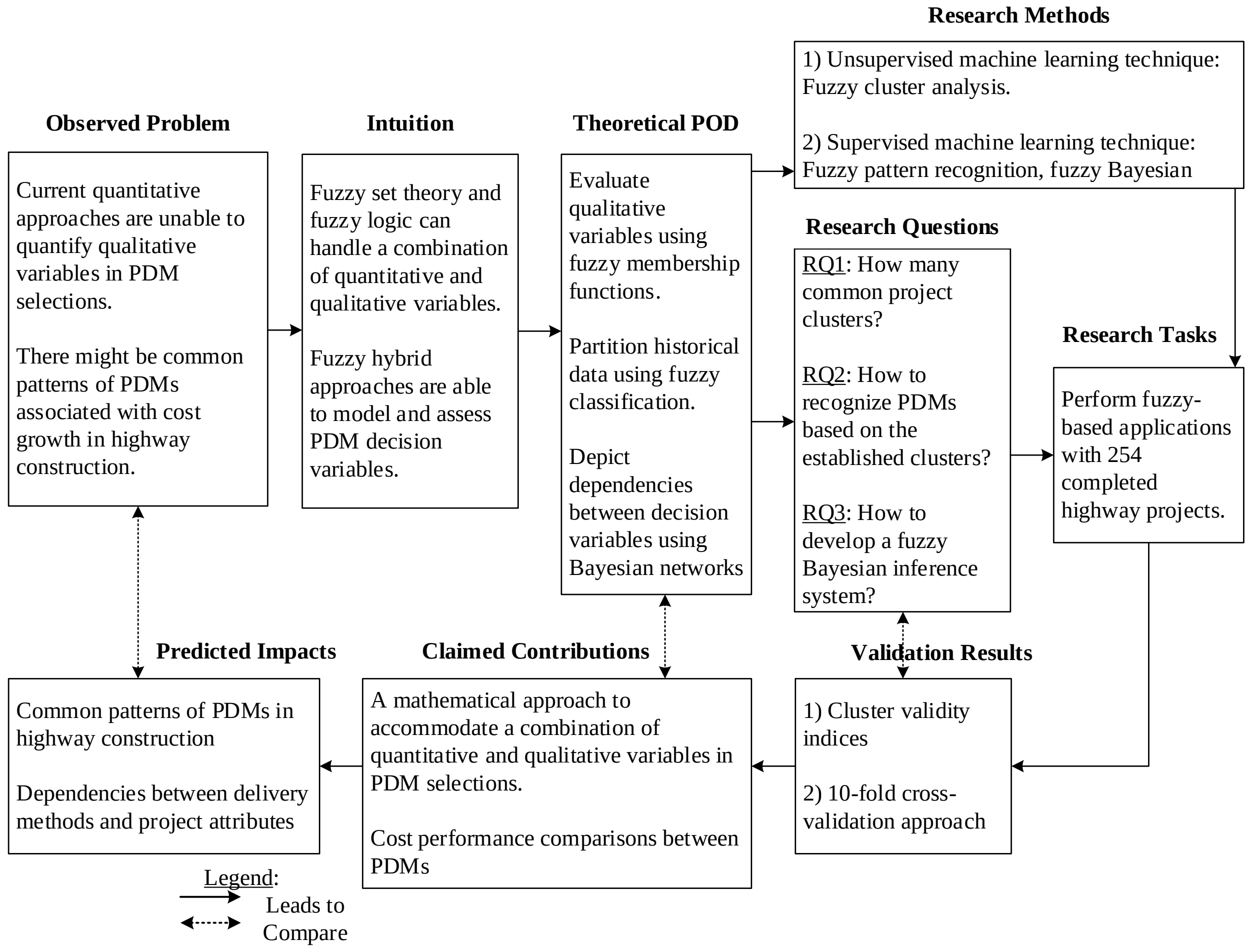

Figure 1. Research Methodology Based on Stanford's CIFE Horseshoe Framework

Research Questions

Research Question 1: Fuzzy Cluster Analysis

What are the underlying clusters of highway projects that share identical project characteristics, cost performance, and the use of the two common project delivery methods (D-B-B and D-B)?

To address the first research question, the point of departure is to implement fuzzy cluster analysis (also known as soft cluster analysis), which is a typical method of grouping highway construction projects in an unsupervised machine learning environment. The performance of all clustering algorithms is dependent on the accurate estimate of the cluster number, which is pre-specified by the analyst. The objective of this technique is to partition a set of *n* highway projects into *C* clusters such that projects within cluster should have similar attributes (i.e., facility types, project types, project complexity, delivery risks, cost growth, and project delivery methods) to each other and vice versa. The results of this question are presented in Chapter 2.

Research Question 2: Fuzzy Pattern Recognition

How do the underlying clusters recognize cost performance's patterns based on inherent project characteristics and delivery methods?

To address the second research question, the point of departure is to implement fuzzy pattern recognition technique, which is a method used to recognize typical cost performance patterns for future projects in a supervised learning environment based on their characteristics and risk profiles. The objective of this technique is to predict cost performance patterns and develop a graphical user interface for the inputs of project characteristics and risk profiles to support decision-making in selecting project delivery methods. The results of this question are presented in Chapter 3.

Research Question 3: Fuzzy Bayesian Inference System

How to develop a Fuzzy Bayesian Inference System based on empirical clusters of highway projects to support selection of project delivery methods?

To address the third research question, a step-by-step fuzzy Bayesian rule-based inference system is developed based on the fuzzy inference system in research question 2. The objective is to demonstrate causality between input variables (i.e., facility type, project type, project complexity, delivery risks) and output variables (project delivery methods associated with cost growth). The results of this question are presented in Chapter 4.

CHAPTER 2:

FUZZY CLASSIFICATION

INTRODUCTION

Selection of the most appropriate delivery method is a key success factor for managing highway construction projects because it can have substantial impacts on project performance, such as cost, time, and quality (U.S. DOT 2006; WSDOT 2016). Comparisons between traditional design-bid-build (D-B-B) and alternative delivery methods, such as design-build (D-B) and construction manager/general contractor (CM/GC), have been investigated for decades. The literature shows there is no single delivery method that is best suited for every construction project; rather, the suitability is practically determined on a project-by-project basis based on unique project characteristics and performance metrics (Minchin et al. 2013; Konchar and Sanvido 1998; Shrestha 2007; Sullivan et al. 2017). Those project attributes are referred to as selection criteria in project delivery decision-making (Touran et al. 2009a).

A typical decision in selecting project delivery methods usually comes from the result of evaluating multiple selection criteria that can be generally split into two main categories: quantitative and qualitative (Gransberg and Shane 2010). Common quantitative selection criteria include project type, size, price, budget, unit cost, cost growth, schedule growth, delivery speed, and construction speed. Common qualitative selection criteria include project complexity, risks, quality, constructability, experience, and innovation. Plenty of statistical and probabilistic approaches have been proposed to quantify project delivery selection criteria and support decision makers (Molenaar and Songer 1998; Tran and Molenaar 2015). Although those quantitative methods perform well in measuring quantitative selection criteria, they are insufficient in measuring qualitative selection criteria. This issue leads to a substantial challenge in project delivery selection because the delivery decision process typically involves a combination of quantitative (e.g., project cost or schedule) and qualitative variables (e.g., project complexity or delivery risk) (Touran et al. 2011).

Current practices of delivery selection in highway construction commonly rely upon experts' experience and guidelines from construction professional organizations (Douglas et al. 2016). These approaches aid project delivery decision-making in the way of studying similar projects and evaluating the most appropriate delivery method in terms of project performance. Identifying commonalities between highway projects in using delivery methods and their performance provides valuable insights and lessons learned for new projects that share similar characteristics. Fuzzy cluster analysis, a data mining technique derived from fuzzy set theory, is capable to explore a set of data and group them based on degree of similarities (Pal and Bezdek 1995). In addition, this approach utilizes fuzzy sets, defined as logical sets whose elements have degrees of membership, to investigate both quantitative and qualitative selection criteria (Ammar et al. 2013).

Motivated by the lack of applications specifically designed to accommodate a combination of quantitative and qualitative variables, which are both critical in selection of project delivery method, the research objectives of this study were to:

- Investigate the latent commonality between highway projects based on project characteristics, cost performance, and project delivery methods used (e.g., D-B-B and D-B);
- Establish clusters of identical highway projects based on the degree of similarity in facility type, project type, project complexity, delivery risks, and cost growth with the labels of D-B-B and D-B; and

- Discuss the appropriate use of D-B-B and D-B for highway projects based on labelled clusters and provide awareness in using historical project data to aid selection of delivery methods.

The rest of this paper is structured as follows. The next section discusses the literature review of the project delivery selection process and applications of fuzzy set theory in construction that derive the research question and the research methodology. These are followed by an illustrative example and discussion. Conclusions are drawn and provided in the final section.

LITERATURE REVIEW

This section discusses the two main areas related to this research: (1) decision-making in project delivery method selection and (2) fuzzy set theory and fuzzy cluster analysis.

Decision-Making in Project Delivery Method Selection

Selection of appropriate delivery methods has been used to improve project performance, including lower cost growth, shorter schedule durations, higher quality, and better safety (Al Khalil 2002; Col Debella and Ries 2006; Ibbs et al. 2003). The traditional D-B-B delivery method is considered to foster adversarial relationships among project participants which often can result in negative performance outcomes (Park and Kwak 2017). On the other hand, alternative contracting methods, including D-B and CM/GC, aim to shorten the project schedule, optimize total cost, and achieve a satisfactory level of project quality (FHWA 2018). Under some particular circumstances, such as projects with a high level of uncertainty or complexity, D-B has been found to provide better project performance than D-B-B (Nikou Goftar et al. 2014; Rojas and Kell 2008). In the U.S., state departments of transportation (DOTs) have increasingly used alternative delivery methods, which inspires the assessment of whether highway projects have achieved better project performance and the identification of common performance patterns to support delivery selection (Touran et al. 2009b).

The majority of current highway project delivery selection approaches are based upon subjective judgements of experts and guidelines from professional organizations in terms of performance of historical project data possessed by the agencies (Bakht and El-Diraby 2015; Mahdi and Alreshaid 2005; WSDOT 2016). Many mathematical and probabilistic decision-making frameworks and models, classified into two main categories: qualitative and quantitative, have been proposed to support project owners in selecting the most suitable delivery method for their project (Al Khalil 2002; Bypaneni 2017; Mostafavi and Karamouz 2010; Tran and Molenaar 2015). There are different types of input variables to project delivery decision-making models, such as project characteristics, project cost and schedule information, project complexity, and delivery risks. As an example of qualitative variables, delivery risks are considered as one of the most difficult inputs to quantify due to the qualitative unit of measurement (Diab et al. 2012). Although delivery risks can be modeled using simulations and probabilistic approaches (Tran et al. 2016), it still preserves a certain level of uncertainty in the model outcomes. It is challenging for statistical and probabilistic models to properly quantify a combination of quantitative and qualitative variables (Al Nahyan et al. 2018).

Fuzzy Set Theory and Fuzzy Cluster Analysis

Fuzzy set theory, a mathematical approach developed by Zadeh (1965) to convert linguistic statements (i.e., qualitative data) to be quantifiable by a computer, have been used to accommodate different combinations of quantitative and qualitative variables (Ammar et al. 2013; Li et al. 2006). The fuzzy set theory has been used in a wide range of domains to evaluate incomplete, imprecise, or qualitative inputs (Chan et al. 2009). In the field of engineering, fuzzy set theory has been used to capture qualitative domain professional judgements to generate theoretical decision-making models and widely applied to many areas, such as computer science, mechanical engineering, aerospace engineering, chemical engineering, and structural engineering (Elwood 2014; D'Urso 2007; Ross 2010; Seo et al. 2004). Within the construction industry, fuzzy set theory has been used in risk-based management (Elbarkouky et al. 2016; Lam et al. 2001; Pawan and Lorterapong 2016).

Derived from fuzzy set theory, fuzzy cluster analysis is used to classify data based on similarities in attributes, features, and other characteristics (Anderberg 2014). This unsupervised learning technique concentrates on grouping data to study underlying data structures and identify the most representative cluster prototypes (Hoppner et al. 1999). There are two common types of cluster analysis: (1) hard cluster analysis, which is developed based on crisp sets, and (2) soft cluster analysis, which is formulated based on fuzzy set theory. Because the scope of this study is to enhance fuzzy set theory to deal with qualitative input data, a soft cluster analysis or fuzzy cluster analysis is conducted. Fuzzy cluster analysis identifies the structure of data and supports establishment of groups of data based on distance with maximum homogeneity (or similarity) within the groups while also having maximum heterogeneity between the groups (Kruse et al. 2007). Fuzzy cluster analysis is different from the crisp cluster analysis in terms of assigning membership values to the clustered data points instead of limiting a single data point to belong to only one cluster (Elwood and Corotis 2015).

In the current literature of project delivery selection, limited research has attempted to provide rigorously means to accommodate a combination of quantitative and qualitative variables and identify groups of projects that share similarities in performance and delivery methods (e.g., D-B-B and D-B) used. To bridge this gap, this study attempted to use fuzzy cluster analysis to group similar highway projects that share high commonalities in project characteristics, project complexity, delivery risks, and cost performance associated with D-B-B and D-B. The clustered groups of similar projects provide insights into recognizing differences in cost performance between D-B-B and D-B highway projects.

RESEARCH QUESTIONS

The main objective of this study was to determine the underlying clusters of highway construction projects based on project characteristics, cost performance, project complexity and risks associated with different delivery methods. To achieve this research objective, this study aimed at addressing the following research questions:

1. What are the latent groupings of cost performance based on delivery methods and degrees of similarity in project characteristics, project complexity, and delivery risks?
2. How do the classified latent groupings differentiate the cost performance of highway projects delivered by D-B-B and D-B?
3. What new information would be gained by using a project delivery-based cost performance cluster analysis?

RESEARCH METHODOLOGY

To address three aforementioned research questions, the research methodology of this study included four steps: (1) literature review; (2) theoretical approach; (3) application; and (4) discussion and conclusion. Figure 1 graphically shows these four steps. First, a comprehensive literature review consisting of two main areas, selection of project delivery methods and the use of fuzzy set theory in construction, was conducted to determine the research gaps. Second, a theoretical approach of how to apply fuzzy cluster analysis to address the identified research gaps was proposed. Third, an illustrative example was presented to discuss the application of the proposed approach. Finally, pairwise comparisons between D-B-B and D-B were discussed. Research contributions, limitations, and future work were summarized in the conclusion section. The following sections briefly discuss the theoretical framework and validation of fuzzy cluster analysis.

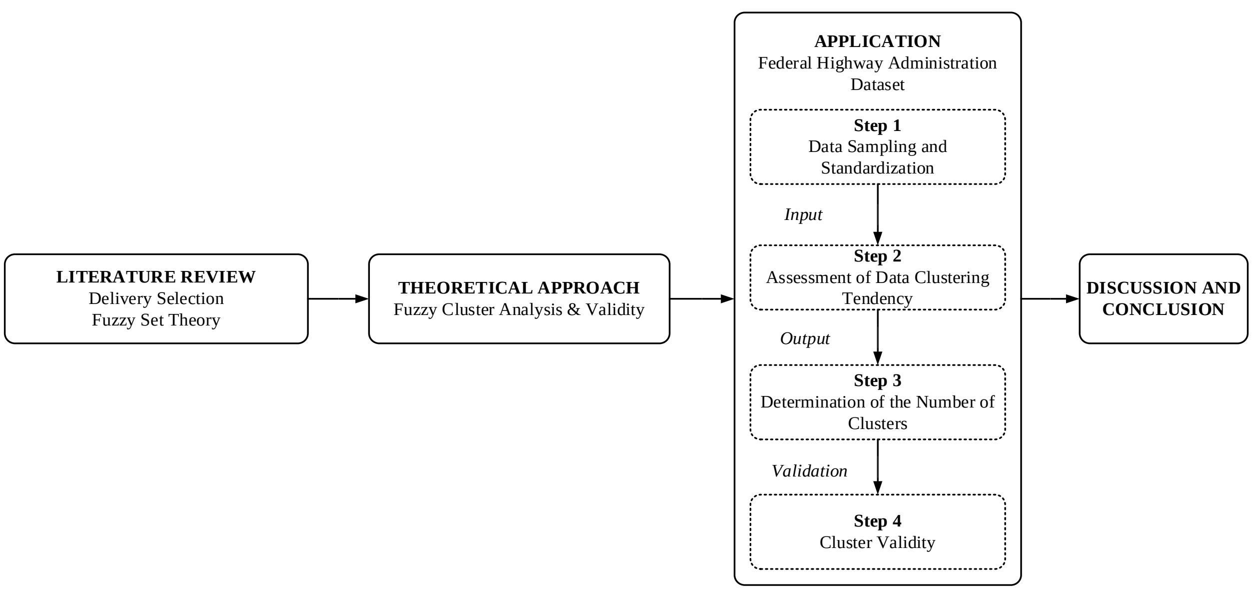

Figure 1. Research methodology

Proposed Approach of Fuzzy Cluster Analysis

Partitioning construction projects into common groups is a complex task that depends on project characteristics, performance, and available data. While fuzzy cluster analysis is commonly used in many research fields, it is relatively new to construction project delivery selection. Building upon from the relevant literature (Elwood and Corotis 2015; Ross 2010; Wu et al. 2010), this study utilized a three-step approach to classifying project characteristics, cost performance, and project complexity and risks associated with D-B-B and D-B delivery methods. These steps are (1) assessment of data clustering tendency, (2) determination of the number of clusters, and (3) validation of the clustering result.

It is noted that the procedure of fuzzy cluster analysis proposed in this study can be implemented in any given set *P* consisting of *n* construction projects, $P = \{P_1, P_2, P_3, ..., P_n\}$. Each project P_i has *m* attributes (e.g., project characteristics and project performance) $P_i = \{P_{i1}, P_{i2}, P_{i3}, P_{i4}, ..., P_{im}\}$. A step-by-step process of this analysis is demonstrated as follow.

Step 1 – Assessment of Data Clustering Tendency

This step is a prerequisite in any clustering processes. There were two methods for assessing the clustering tendency used in this study: (1) a statistical method (Hopkins statistics) and (2) a visual assessment of cluster tendency (VAT) algorithm. The Hopkins statistic method measures the probability with which the dataset *P* is established by a uniform data distribution to examine the spatial randomness of the dataset. This method first iteratively calculates the distance between a project P_n and its nearest neighbor P_{n+1}, denoted as X_i. Then, it generates a simulated dataset *S* drawn from the given dataset *P* and iteratively computes the distance between a simulated project S_n and its nearest neighbor S_{n+1}, denoted as Y_i. The Hopkins statistic (H value) is then the ratio of the average nearest neighbor distance in the simulated dataset S to the sum of the average nearest neighbor distances in both datasets P and S as shown in Eq. (1) (Kassambara 2017). If the H value is equal or greater than 0.5, collected projects in the construction dataset P are uniformly distributed, which does not give any underlying groups.

$$H = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i} \tag{1}$$

The second method of assessment, VAT algorithm, provides a graphical determination for the clustering tendency by first calculating the dissimilarity matrix between projects in the dataset P and then reordering the calculated matrix to make sure that similar projects are close to one another. The ordered matrix is visualized in images as outputs of the VAT method to confirm whether there is a clustering structure in the dataset P.

Step 2 – Determination of the Number of Clusters

If the construction dataset P is eligible to utilize fuzzy cluster analysis, the next task is to determine an appropriate clustering algorithm to optimize the goodness of the final clustering groups that logically represent the latent structures of clustering data. Fuzzy C-means algorithm (FCM), the most common soft clustering method in the domain of unsupervised machine learning techniques (Elwood and Corotis 2015), was used in this study. This algorithm takes into account any combinations of both quantitative and qualitative variables. The FCM assigns the projects into clusters based on the degree of fuzzy membership $\mu_{ij}(x) = \mu_C(P_i) \in [0,1]$, where μ_{ij}

is the value that describes the degree of membership of project P_i in the j^{th} fuzzy cluster $C_j = \{1, 2, 3, ..., k\}$ with k is the number of clusters. Using this algorithm provides two main outcomes: a membership matrix U and a vector of cluster centers v_{ij}. The membership matrix describes the degree of membership of each project within the identified cluster while the vector of cluster centers represents the features of the identified clusters as calculated by Eq. (2). These two outcomes help determine an appropriate number of clusters in the dataset P by grouping projects P_i having similar attributes and membership values.

$$v_{ij} = \frac{\sum_{j=1}^{n} \left[\left(\mu_{ij} \right)^m \times P_{ij} \right]}{\sum_{j=1}^{n} \left(\mu_{ij} \right)^m} \tag{2}$$

The number of underlying clusters within the construction dataset P is determined by using the optimization function J_m, which attempts to simultaneously minimize the distance between projects within a cluster and maximize the distance between clusters as shown in Eq. (3). U is the membership matrix; v is the vector of cluster centers; and d_{ik} is the distance between the cluster center v_i and its assigned projects P_i as shown in Eq. (4). When the assignment of projects to particular clusters reaches to the minimum of J_m, the final number of underlying clusters within the dataset P is recognized.

$$J_m(U, v) = \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\mu_{ij} \right)^m \left(d_{ij} \right)^2 \tag{3}$$

$$d_{ij} = d\left(P_i - v_j \right) = \left[\sum_{j=1}^{m} \left(P_{ij} - v_{ij} \right)^2 \right]^{1/2} \tag{4}$$

Step 3 – Validation of the Clustering Result

To validate the goodness of the clustering result for the given dataset *P*, this study proposed four typical clustering validity indices: partition entropy, partition coefficient, the Dunn index, and the silhouette width. Each index evaluates the degree of decency between each project in the dataset *P* and the identified clusters in terms of within- and between-cluster agreements. These indices examine if the classified projects are in a proper cluster. Using cluster validity helps avoid randomness in identifying clusters to provide better recognition of underlying structures within the dataset. Since fuzzy cluster analysis is one of unsupervised learning techniques, which concentrates more on exploration of data structures, it does not require any comprehensive validation processes (Ross 2010). The determination of the optimal number of clusters, thus, often depends on subjective judgements of the analyst (Elwood 2014). The following section exemplifies the proposed approach in detail though discussing an illustrative example of selecting delivery methods in highway construction projects.

APPLICATION

This section provides an illustrative example of applying fuzzy cluster analysis to determine the underlying clusters of highway construction projects based on project characteristics, cost performance, and project complexity and risks associated with D-B-B and D-B delivery methods. All analyses were conducted using an R programming environment with multiple relevant clustering packages. As shown in Figure 1, data preparation via sampling and standardizing processes was presented first before illustrating the three-step procedure of fuzzy cluster analysis.

Data Sampling

This study utilized an empirical construction dataset of 254 highway projects collected from 28 state DOTs. More information of the data collection process and analysis results can be found from FHWA (2018). Collected projects consisted of various characteristics, including facility type, project type, complexity rating, cost performance, project delivery methods, and risk rating. Initially, the collected dataset had 291 projects, but 37 projects were eliminated due to missing data and extreme outliers. Fuzzy cluster analysis is critically influenced by outliers (Kruse et al. 2007). Thus, identified outliers from the dataset were removed. Prior to conducting cluster analysis, selection of major input variables was required. Five main project characteristics, including facility type, project type, complexity, risk profiles, and delivery types, and project cost performance were selected from the survey instruments. Each project characteristic has a particular number of sub-categories as shown in Table 1. Accordingly, 17 variables were used as the features of the subsequent clustering process, consisting of 16 sub-categories from the five selected project characteristics and the project cost growth. In addition, project delivery (D-B-B and D-B) was used as a control variable in this analysis.

Table 1. Selected features of the fuzzy clustering process (m = 17)

Facility Type ($m_1 = 5$)	**Project Type ($m_2 = 3$)**	**Project Complexity ($m_3 = 1$)**	**Risk ($m_4 = 7$)**	**Cost Performance ($m_5 = 1$)**
Road	New construction	Complexity	Complexity Risks	Cost growth
Bridge	Reconstruction		Quality Risks	
Drainage	Other		Constructability Risks	
ITS*			Construction Risks	
Other			Utility and right-of-way Risks	
			Management Risks	
			Environmental Risks	

*Note. *ITS: intelligent transportation system*

Facility Type

This variable consists of five sub-facility types: road, bridge, drainage, and intelligent transportation system (ITS) based on approximate percentages of the total project cost. The range of each type is from 0 to 100%. For example, a highway project might consist of a scope that is 80% road, 10% bridge, and 10% ITS; the total should be always 100%. Facility type contains continuous data and provides five input variables.

Project Type

This variable includes three sub-project types: new construction, and reconstruction based on approximate percentages of the total project cost. Similar to facility type, the range of each type is from 0 to 100%, and it follows the continuous data type. For instance, a highway project might have two-thirds of new construction and one-third of resurfacing. Project type provides three input variables.

Project Complexity

This variable is rated based on a 3-point ordinal scale. First, the "most complex" projects include those that are new highways and major relocations, new interchanges, capacity adding and major widening, major reconstruction, require congestion management studies, and have complex environmental assessment or environmental impact statements. Second, the "moderately complex" projects include those that are minor roadway relocations, non-complex bridge replacements with minor roadway approach work, and non-complex environmental assessment required. Third, the "non-complex" projects include those that are maintenance betterment projects, overlay projects with simple widening, little or no utility coordination, non-complex enhancement projects without new bridges, and categorical exclusion.

Project Risk

This variable consists of thirty-one project risks rated with a 5-point ordinal scale as shown in Figure 2, which represents five potential degrees of risk impacts (very low, low, moderate, high, and very high) on cost performance as known by the project team before the beginning of the project to denote the project's level of riskiness. Because of a large number of project risks, it is not sufficient to use them all for cluster analysis. Therefore, this study observed seven risk factors as the final result of using exploratory factor analysis (EFA). Since the discussion of the EFA process is beyond the scope of this study, details can be found in Bypaneni (2017). The resulting seven risk factors from EFA were:

- Risk factor 1, "complexity risk", consists of five risks: project complexity, uncertainty in geotechnical investigation, legal challenges, and changes in law, intergovernmental agreements and jurisdiction, and difficulty in obtaining other agencies.
- Risk factor 2, "quality risk", consists of two risks: construction quality control and quality assessment process and design quality assurance.
- Risk factor 3, "constructability risk", consists of two risks: delays in procuring critical materials, labor, and specialized equipment and significant increase in material, labor and equipment cost.
- Risk factor 4, "construction risk", consists of two risks: work zone traffic control and construction sequencing, staging, and phasing.
- Risk factor 5, "utility and right-of-way (ROW) risk", consists of three risks: unexpected utility encounter, delays in completing utility agreements, and delays in ROW process.
- Risk factor 6, "management risk", consists of three risks: staff experience and availability, project and program management issues, and conformance with regulations, guidelines, and design criteria.

- Risk factor 7, “environmental risk”, consists of two risks: challenges to obtain appropriate environmental documentation and environmental impacts. There are seven input variables from risk factors.

PROJECT RISK PROFILE

Rating System	1	2	3	4	5
	Very Low	*Low*	*Moderate*	*High*	*Very High*
Cost Impact	Insignificant cost increase	< 2% cost increase	2-5% cost increase	5-10% cost increase	> 10% cost increase

Note: NA (Not Applicable) is used to rate risk that is not materialized.

Risk Description	Rating					
1. **Challenges to obtain appropriate environmental documentation** — changing environmental regulations, unforeseen formal NEPA consultation, unexpected Section 106 issues, an insufficient environmental study, environmental clearance for staging required, etc.	NA	1	2	3	4	5
2. **Environmental impacts** — unexpected environmental constraints during planning and construction (e.g., historic site, endangered species, wetland, coastal and scenic zone, and wildlife; Environmental Assessment vs. Environmental Impact Statement).	NA	1	2	3	4	5

Figure 2. Example of data collection form

Cost Performance

Cost performance consists of various construction cost data, such as engineer’s estimate, contract award, and final cost. The final cost is equal to the contract award plus costs of all change orders. To represent the cost performance variable, this study used cost growth, which is the overall performance at project completion, calculated from the project cost data using Eq. (5).

$$\textit{Cost Growth (\%)} = \textit{(Final Cost} - \textit{Contract Award)(100\%)/(Contract Award)} \qquad (5)$$

Table 2 provides descriptive statistics of the cost growth variable. Cost growth values were divided into five separate qualitative groups of “saving”, “none”, “low”, “medium”, and “high.” These five groups are supported by cost-performance studies in the construction literature (Chen et al. 2016; Love et al. 2013; Sullivan et al. 2017). The majority of the cost growth values fall within the range from -1% to 5%.

Table 2. Descriptive statistics of cost performance (n = 254)

Cost Growth Group	Size (n)	Range (%)	Mean (%)	Standard Deviation
Saving	48	-10 to -1	-4.43	0.029
None	60	-1 to 1	0.11	0.004
Low	70	1 to 5	2.98	0.012
Medium	36	5 to 10	7.02	0.011
High	40	10 to 20	15.12	0.038

Data Standardization

The collected FHWA data with 17 selected variables were standardized prior to conducting fuzzy cluster analysis because the selected variables had different types of measuring unit. To assess the similarity of two projects, fuzzy cluster analysis calculates the Euclidean distance between them, which is a geometric measure of closeness between data points (Kassambara 2017). The value of this distance is closely related to the measuring scale of selected variables and influences the shape of the clusters. Thus, there should be a unified scale within the dataset to avoid impacts of dissimilar measures; for instance, there is no valid comparison between variables “project type” and “risk rating”. Within the domain of fuzzy cluster analysis, ranging is one of the recommended methods to standardize data based on the max and min values of the input variables (Elwood 2014; Kassambara 2017; Kruse 2007). The ranging method was conducted separately for each variable. In other words, the data standardization of “facility type” did not affect the standardization of “project type”, “complexity”, or “risk factor”. Each variable was treated specifically based on its attribute sample and own range to preserve the relationships among data points, according to previous studies (Elwood and Corotis 2015). This process ensures that the numerical and categorical variables were properly standardized to use for fuzzy cluster analysis. The standardization process resulted in 17 standardized variables with the same scale from 0 to 1.

Assessment of Data Clustering Tendency

Investigation of clustering tendency assessment of the FHWA dataset, with two methods: Hopkins statistic and VAT, confirmed that this highway dataset contained several latent groups in terms of cost growth. The Hopkins method used in this study examined two hypotheses:

- Null hypothesis: the FHWA dataset is uniformly distributed which indicates that no distinct grouping of cost performance exists in this dataset.
- Alternative hypothesis: the FHWA dataset is not uniformly distributed which indicates that potentially distinct groupings of cost performance exist in this dataset.

Using Eq. (1) with R functions provided a Hopkins statistic value of 0.274 that was far below the threshold of 0.5; as a result, the FHWA dataset was clusterable (Kassambara 2017). Based on the VAT method, the clustering tendency was visually assessed by graphically counting the amount of dark squares along the diagonal of a dissimilarity matrix in the VAT image (Figure 3). It is noted that the red squares in Figure 3 illustrate data with high similarity (values close to 0). The purple squares illustrate data with low similarity (values away from 0). The dissimilarity matrix image in this figure clearly shows the different areas of similar and dissimilar projects indicating that the data can be clustered. Accordingly, the VAT method was also in line with the Hopkins method to confirm that there was a substantial clustering structure in the FHWA dataset. This implies that there is a variation in terms of cost performance between the use of D-B-B and D-B in highway construction projects.

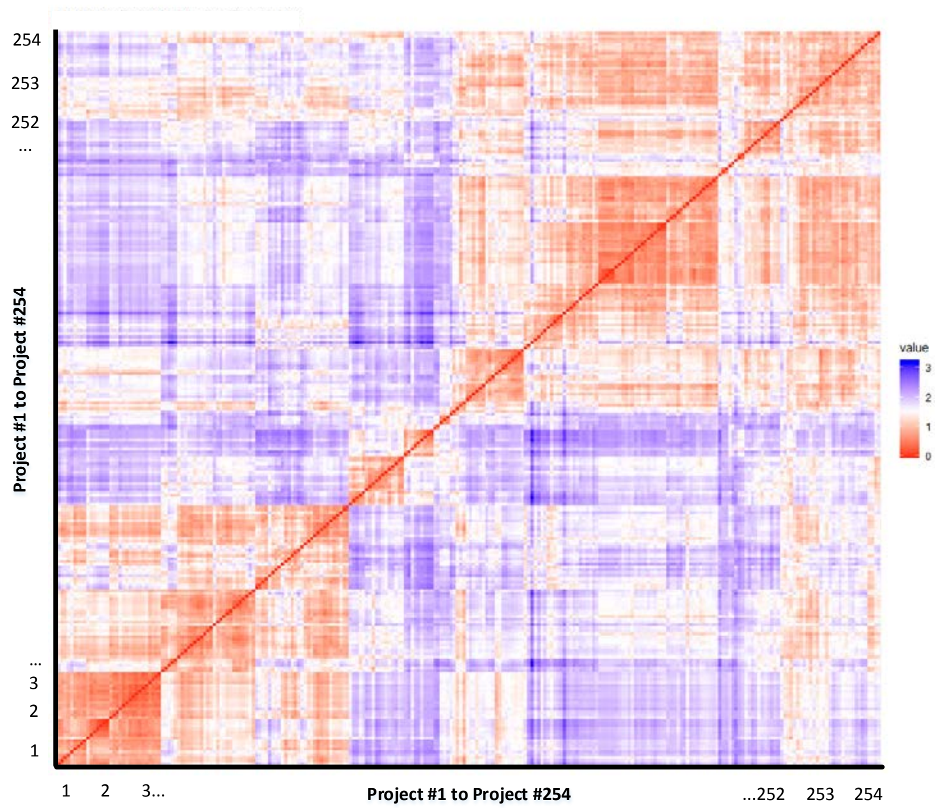

Figure 3. Visual assessment of cluster tendency (VAT) algorithm

Determination of Number of Clusters

When using FCM, an individual project is assigned a degree of membership in each cluster based on the similarity of cost performance. Accordingly, a project can belong to multiple clusters, which shows the overlapping characteristic of fuzzy clustering groups. The determination of the optimal number of clusters, thus, turns into a complicated procedure. If the number of clusters is too small, the distance between clusters becomes wider (i.e., the variation between clusters is high); on the other hand, if this number is too large, the distance between projects becomes smaller (i.e., the clustering result is less informative).

To identify an appropriate number of common groups of cost growth within the FHWA dataset, this study used four methods, including visualization-based (Elbow and Silhouette methods) and statistics-based (gap statistics and NbClust methods). All four methods utilized the FCM algorithm to compute and compare possible alternatives of the optimal number of clusters that could be extracted from the FHWA dataset, ranging from two to fourteen clusters. Each alternative contains a membership matrix U and a vector of cluster centers v_{ij} of 254 highway projects calculated based on Eq. (2). To compare the alternatives, the four methods iteratively computed and minimized the optimization function J_m of each alternative based on Eq. (3). The computing process in this study was handled by R functions defined for fuzzy cluster analysis.

Based on visualization, the Elbow method selects a number of clusters from which adding another cluster does not increase the total within sum of squares (WSS), measured by the total distances between assigned projects. The Silhouette method evaluates how well each project lies within the associated cluster. The optimal number of clusters is the one that achieves the highest value of the average silhouette. Based on statistics, the gap statistics method produces the optimal number of clusters by comparisons of the total within intra-cluster variation for different numbers of clusters with statistically anticipated values. The maximum gap statistic value provides the optimal number of clusters. The NbClust method examines more than thirty clustering indices to select the optimal number of clusters based upon the majority rule. Table 3 summaries detailed information and selection criteria of the four methods to identify the most optimal number of clusters. The results from the four methods indicated that the range of the potential numbers of clusters was from two to fourteen clusters. Among the recommended numbers, the option of seven clusters produced the lowest WSS (248.48) in the Elbow method, the highest average silhouette value of 0.25, the highest gap statistic value of 0.59, and the majority of clustering indices (n=8). Thus, the number of seven clusters was determined as the most optimal number of clusters that could be extracted from the given dataset. Accordingly, seven clusters were defined as the input for the number of cluster centers to the FCM algorithm.

Table 3. Determination of the optimal number of clusters for cost performance

Method	Elbow	Silhouette	Gap Statistics	NbClust
Mechanism	Determines the total within sum of square (WSS) which measures the clustering compactness	Examines how well data points are clustered	Compares WSS to assess if clustering results are far away from the uniform distribution	A pre-defined function in R programming which provides 30 indices for identifying the optimal number of clusters
Advantage	Easy to use	Can be used for any clustering technique	Can be used for any clustering technique	Highly reliable
Disadvantage	Only for c-means clustering technique	Only for global clustering attributes	Complex algorithm	Only used with R packages
Selection Criteria	Minimum value of WSS	High average silhouette value	Maximum value of the gap statistics	Majority rule for 30 clustering indices

It is noted that the FCM algorithm first identified seven cluster centers, and then assigned data points to the appropriate clusters based on the closeness of the data points to the cluster centers by calculating the distances between them. Essentially, this algorithm concurrently minimized the distance between assigned projects within a cluster and maximized the distance between seven clusters based on similarity and dissimilarity in five selected features: facility type, project type, project complexity, project risk, and cost growth. As a result of using FCM, seven clusters experienced overlapping based on degrees of memberships of each project to specific clusters (Figure 4). It is important to note that Figure 4 shows the separation of the seven clusters grouped from 254 highway projects associated with a 2-D illustration of a 17-D space. A detailed process of using FCM to minimize the within-cluster distances and maximize the between-cluster distances is provided in Ross (2010). After clustering, the partitioned projects were hardened into crisp clustering groups for subsequent humanistic judgements. This process is called defuzzification, where projects obtained the maximum membership value within a cluster would be assigned to that cluster. This helps reduce fuzzy information and enhance the degree of interpretation of fuzzy-based results.

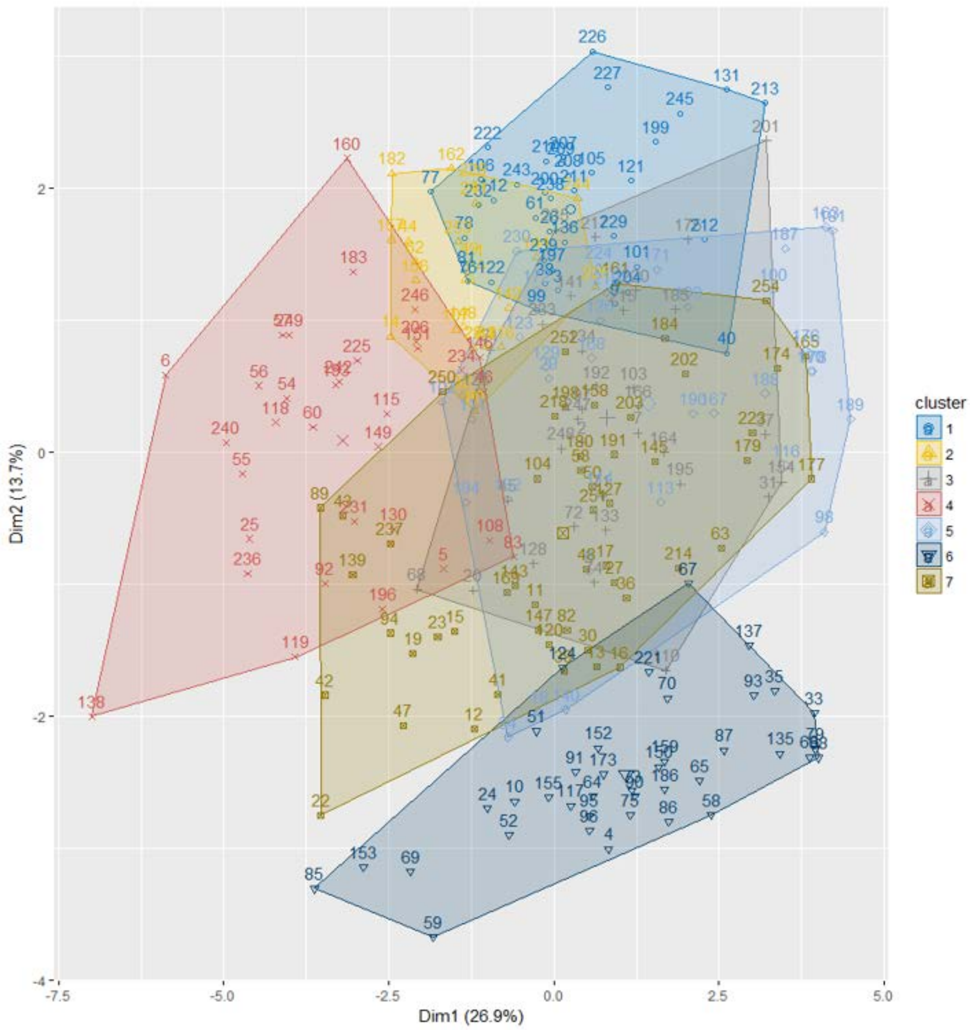

Figure 4. Graphical demonstration of seven clusters

Table 4 summarizes the classified seven clusters associated with dominant features. According to fuzzy set theory, the selected features were assigned to each cluster based on their memberships within clusters (Wu et al. 2010). The cut-off point of 20% was used to determine representatives of each cluster. Cluster 1 represents saving-low cost growth, moderately complex, reconstruction road projects with very low to low risk impacts. Cluster 2 represents low cost growth, moderately complex, new construction bridge projects with very low to low risk impacts. Cluster 3 represents medium-high cost growth, the most complex, new construction, road projects with very low to low risk impacts. Cluster 4 represents saving-low cost growth, the most complex, new construction road projects with low risk impacts. Cluster 5 represents low cost growth, moderately complex, road projects with other project type and very low risk impacts. Cluster 6 represents none cost growth, moderately complex, reconstruction-bridge projects with very low risk impacts. Cluster 7 represents none to low cost growth, the most complex, new construction road or bridge projects with low to medium risk impacts.

Table 4. Characteristics of seven classified clusters of cost performance (n=254)

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Size (n)	52	34	39	34	23	39	33
Delivery Method	D-B-B	D-B	D-B-B	D-B	D-B	D-B-B or D-B	D-B
Cost Growth	Saving-Low	Low	Medium-High	Saving-Low	Low	None	None-Low
Dominant Facility Type	Road	Bridge	Road	Road	Road	Bridge	Road or Bridge
Dominant Project Type*	Recon	New	New	New	Other	Recon	New
Complexity	Moderate	Moderate	Most	Most	Moderate	Moderate	Most
RF 1 – Complexity Risks**	VL***	VL	L	L	VL	VL	L
RF 2 – Quality Risks	VL	VL	VL	L	VL	VL	L
RF 3 – Constructability Risks	VL	VL	VL	VL	VL	VL	L
RF 4 – Construction Risks	L	L	L	L	L	VL	M
RF 5 – Utility and ROW Risks	VL	VL	L	L	VL	VL	L
RF 6 – Management Risks	VL	VL	VL	VL	VL	VL	L
RF7 – Environmental Risks	L	L	L	L	VL	VL	M

Note. *Reconstruction and new construction; **RF: Risk Factor; *** VL – very low; L – Low; M – medium; H – high; VH – very high

Cluster Validity

Cluster validation refers to assessing of the goodness of clustering results. Specifically, the decency of determination of the optimal number of clusters is important to ensure the reliability of clustered results (Elwood 2014). This study employed four common cluster validity indices in the domain of fuzzy cluster analysis as shown in Table 5. The partition entropy index should be minimized while the remaining indices (partition coefficient, silhouette width, and Dunn index) should be maximized to have more accurate clustering results (Říhová and Makhalova 2017). The results range from “fair” to “good”, which satisfies the validation of cluster analysis (Kassambara 2017; Pal and Bezdek 1995; Wu and Yang 2005). The cluster validity indices are used to avoid finding clusters in noise, select an appropriate clustering algorithm, and compare the identified clusters. This study employed the four cluster validity indices to check if the projects in the same cluster were similar as much as possible and the projects in different clusters were highly distinct. In addition, the clusters generated in this study reasonably match with labels of delivery methods (D-B-B and D-B) and cost growth. Therefore, the “fair” and “good” results indicate that the seven identified clusters truly satisfy the requirement of cluster validity. It is more important that the result of seven clusters can be meaningfully labeled in terms of project delivery methods (i.e., D-B-B and D-B) to produce subsequent comparisons. This yields that fuzzy cluster analysis can be used to classify different groups of cost performance of highway projects delivered by D-B-B and D-B.

Table 5. Cluster validity indices

Index	Partition Entropy	Partition Coefficient	Silhouette coefficient	Dunn Index
Definition	Represents the fuzziness in clusters	Measures overlaps among clusters	Estimates the average distance between clusters and data points within a cluster	Measures compactness of clusters and distance between clusters
Advantage	Easy to use	Highly reliable	Easy to use	Can be used for any clustering algorithm
Disadvantage	Sensitive to outliers	Complex algorithm	Sometime ambiguous	Computational effort
Decision Criteria (from 0 to 1)	0 – Excellent 0 to 0.5 – Good 0.5 to 1 – Fair 1 – Bad	0 – Bad 0 to 0.5 – Fair 0.5 to 1 – Good 1 – Excellent	0 – Bad 0 to 0.5 – Fair 0.5 to 1 – Good 1 – Excellent	0 – Bad 0 to 0.5 – Fair 0.5 to 1 – Good 1 – Excellent
Validity Result	0.46	0.11	0.58	0.11

DISCUSSION

Using fuzzy cluster analysis, this study addressed the research question 1 by identifying seven groupings of cost performance based on the degree of similarity in project characteristics, project complexity, project risks, and the use of different delivery methods as summarized in Table 4. With different selected project characteristics, the majority of the common groups of cost growth in the collected highway projects scattered to "none" and "low" while one group represented the "medium to high" cost growth. Based on these identified common groupings, this study concentrated on discussion of differences between D-B-B and D-B highway projects.

Discussion of pairwise comparisons of cost performance in D-B-B and D-B highway projects helps address research questions 2 and 3. The results of this study imply that D-B generally outperformed D-B-B in terms of cost performance, which is consistent with previous studies (CII 2018; Goodrum et al. 2011; Hale et al. 2009; Shrestha 2007). For instance, Shrestha et al. (2007) investigated empirical transportation data from Texas DOT to conclude that D-B projects achieved superior cost performance than D-B-B projects. However, results of the present study are not in line with Minchin et al. (2013), which compared sixty highway and bridge projects from Florida DOT and stated that using D-B-B resulted in better cost performance than using D-B. A possible reason is that this study takes into account uncertainty impacts of project complexity and risks while Minchin et al. (2013) did not include these inherent factors. According to Table 4, in the same level of project complexity, D-B bridge projects showed less cost growth and lower risk impacts on cost performance than D-B road projects. On the other hand, with the use of D-B-B, road-typed projects, which suffered higher cost impacts of seven risk factors, had higher cost growth than bridge-typed projects. The following sections demonstrate detailed comparisons of D-B-B and D-B in road, bridge, new construction, and reconstruction projects based upon the results of fuzzy cluster analysis.

Road Projects

Table 4 shows that road projects delivered by D-B performed better than D-B-B in terms of cost growth. This result is in line with Shrestha et al. (2012) and Tran et al. (2018). According to cluster 1, D-B-B showed a low cost growth (mean = 5%) in reconstruction road projects with low complexity and very low risk impacts. On the other hand, D-B in cluster 4 showed a low cost growth (mean = 2%) in new construction road projects with high complexity and risk impacts. In new construction roads (clusters 3, 4, and 7), projects procured by D-B showed a lower cost growth than projects delivered by D-B-B even though D-B projects had higher complexity and risk impacts. For instance, in cluster 7, D-B showed virtually no cost growth (mean = 0%) in new construction road projects despite having a higher level of risk impacts on cost performance compared to cluster 3 where D-B-B showed a high cost growth (mean = 15%) in the same project type.

Bridge Projects

D-B-B reconstruction bridge projects achieved a lower cost growth than D-B delivered new construction bridge projects according to clusters 2, 6, and 7 (Table 4). This is not in line with several project delivery studies, such as Touran et al. (2011) and Sullivan et al. (2017). In their study analyzing nine case study transit projects in the U.S., Touran et al. (2011) descriptively concluded that D-B produces more cost savings than D-B-B; additionally, no statistical inference was found in their study. According to the comprehensive literature review study of Sullivan et al. (2017), D-B-B showed a higher cost growth (mean = 5.1%) compared to D-B (mean = 2.8%) in an investigation of 2,919 construction projects. These studies, however, did not consider project complexity and risks in the comparison between D-B-B and D-B. This study, on the other hand, takes into account these two attributes in the clustering process. Since these two project attributes contribute significantly to the variation of the cost performance (Touran et al. 2009a; Tran et al. 2016), this may explain the unexpected comparison results between the two delivery methods. For example, the D-B-B bridge projects in cluster 6 were reconstructed with a low level of complexity and very low risk impacts, making the corresponding result of very low cost growth (mean = 0%) appear to be logical.

Project Complexity and Risks

In highway projects with a higher level of complexity and greater risk exposure, this study generally found that D-B outperformed D-B-B. From clusters 4 and 7, D-B showed a low cost growth (ranging from 0% to 5%) in highway projects characterized with the highest level of complexity and risk impacts within the FHWA dataset. In addition, the construction and environmental risk factors had a high impact on D-B project cost performance. On the other hand, D-B-B, in cluster 3, produced a high cost growth (mean = 15%) in highway projects with the same level of complexity and even lower levels of risk impacts. This finding supports the use of D-B in complex highway projects and is in line with Park and Kwak (2017) in which D-B is recommended to use for transit projects that have a high level of complexity and involve a large number of project risks.

CONCLUSIONS

This study investigates the similarity in cost performance between highway projects based on project characteristics, project complexity, and delivery risks to identify the pattern of project performance delivered under D-B-B and D-B. Since the decision-making process of delivery selection depends upon a combination of quantitative and qualitative variables, this study implements fuzzy set theory in the context of fuzzy cluster analysis, which is an effective approach for simultaneously considering quantitative and qualitative variables. The proposed approach provides the utilization of historical project data available in construction to support project delivery decision-making.

The findings of this study successfully addressed the three research questions. For the research question 1, as one of the first attempts to implement fuzzy cluster analysis in the domain of delivery selection, this study identified seven common groups of cost performance based on facility type, project type, project complexity, and delivery risks under the use of D-B-B and D-B. For the research questions 2 and 3, the pairwise comparisons between D-B-B and D-B derived from the seven identified clusters provide insights into cost performance of these two delivery methods in the highway sector. D-B-B produces a large variation of cost growth (ranging from 0% to 20%) in highway projects. Specifically, it produces low cost growth (mean = 5%) for cluster 1, but medium to high cost growth (mean = 15%) for cluster 3. On the other hand, D-B shows a relatively small variation of cost growth (ranging from -1% to 5%) in highway projects. Particularly, it produces low cost growth (mean = 2.25%) for clusters 2, 4, 5, and 7. For cluster 6, there is no difference in cost growth between D-B-B and D-B. This study also found that D-B typically outperformed D-B-B in new construction, complex, and highly risk-involved projects whereas D-B-B was a better choice in certain reconstruction projects in the highway sector.

Research Contributions

This study contributes to the project delivery body of knowledge by identifying seven clusters of highway projects that share high commonalities in project characteristics, project complexity, delivery risks, and cost performance associated with the use of D-B-B and D-B. The proposed application is a rigorous solution for measuring a combination of quantitative and qualitative variables to support decision-making in project delivery selection. This application can be applied to other decision scenarios in construction that require both quantitative and qualitative inputs. Within the decision scenario context of delivery method selection, the proposed approach, fuzzy cluster analysis, recognized distinct clusters that enables the pairwise comparisons between them. With the common groups of cost performance identified based on D-B-B and D-B, this study confirms the applicability of fuzzy cluster analysis in the domain of project delivery selection.

To the construction industry, the seven identified clusters of similar highway projects help practitioners identify empirical trends of cost performance under the use of D-B-B and D-B. These common groupings of cost performance may aid DOT's agencies in making initial project delivery decisions in the feasibility and planning phases of new highway projects where very little information was available. This study also provides different scenarios of comparing D-B-B and D-B in the highway sector based on facility type, project type, project complexity, delivery risks, and cost performance identified in seven clusters. The finding of this study provides a valuable decision aid because the decision maker can match and compare a new project's inputs with the features of the identified clusters to determine an appropriate delivery method under different scenarios.

Limitations and Recommendations for Future Research

There are several limitations in this study. First, this study only conducted fuzzy cluster analysis for D-B-B and D-B projects because of the limited amount of CM/GC highway project data available. Future research may need to collect more completed CM/GC highway projects to overcome this limitation. The second limitation of this study is that only highway construction projects were considered. Different project types from the vertical sector or other infrastructure projects can benefit from similar analysis. Further, additional project attributes, including project size, procurement method, and payment method may be incorporated. Finally, this study did not concentrate on evaluating the extent to which the various input factors were linked with cost performance outcomes; rather, the intent of this exploratory study was to essentially identify different groupings of cost performance within highway projects delivered by D-B-B and D-B based on the degree of similarity of five particular project characteristics. Therefore, the varying effects between selected project characteristics and cost performance were not examined, and this is considered an opportunity for future work. In addition, future studies may utilize additional applications of fuzzy set theory in the environment of supervised learning techniques, such as fuzzy logic, fuzzy inference system, and fuzzy pattern recognition, to develop a comprehensive decision-aid system for selection of delivery methods.

CHAPTER 3:

FUZZY PATTERN RECOGNITION

INTRODUCTION

The construction industry contains a large amount of historical data that can be used to analyze potential patterns between project delivery methods, project performance, and various project attributes. Since there is no "one-size-fits-all" delivery method for construction projects, project delivery selection is a critical task in early pre-construction decision-making processes that relies on many criteria (Shrestha et al. 2012; Sullivan et al. 2017; Touran et al. 2009; Tran et al. 2013). Many delivery selection decision-aids have been proposed for decades to help select an appropriate delivery method from the traditional design-bid-build (D-B-B) to alternative delivery methods, such as design-build (D-B) and construction manager/general contractor (CM/GC). The previously developed decision-aids provide a robust process to quantify historical project data and investigate project performance (e.g., cost, schedule, and quality) associated with different delivery methods (Tran et al. 2015).

Project delivery selection often involves a variety of qualitative data in terms of subjective judgements and opinions of experts in addition to quantifiable data (Douglas et al. 2016). For instance, risk assessment and analysis play a pivotal role in determining an appropriate delivery method (Tran and Molenaar 2015). Current scientific and quantitative approaches enhance statistical probabilities to robustly assess project performance based on historical data. For example, several researchers have developed decision-aided models based on probabilistic approaches (Molenaar and Songer 1998; Tran and Molenaar 2015) for selecting an appropriate delivery method. However, there still exists a challenge in the decision scenarios where a combination of quantitative (i.e. numeric) and qualitative (i.e. categorical) inputs should be sufficiently accommodated (El Asmar et al. 2013). The currently deployed probabilistic approaches are not an ideal means to address the uncertainty encountered in qualitative data inputs. Fuzzy set theory, on the other hand, is a decision-aid tool that is efficient in modeling subjective expressions and investigating a combination of quantitative and qualitative data (Ross 2010). Few studies have attempted to apply statistical techniques with fuzzy sets to develop sophisticated hybrid models that are specifically suited to accommodate both quantitative and qualitative inputs in delivery selection (Mafakheri et al. 2007; Mostafavi and Karamouz 2010).

The objective of this study was to investigate the applicability of fuzzy pattern recognition for delivery selection for highway design and construction projects. Specifically, this study aimed at investigating how to recognize potential delivery methods associated with cost growth for new highway projects by using common delivery patterns built upon historical data. To achieve the research objective, this study has attempted to complete the following research tasks:

1. Select and build fuzzy membership functions of facility type, project type, project complexity, delivery risk, and cost performance under the use of different delivery methods (e.g., D-B-B and D-B);
2. Develop a fuzzy rule-based inference system based on particular project attributes and a collected empirical highway dataset using R programming;
3. Demonstrate the process of fuzzy pattern recognition for highway construction projects through a case example; and
4. Discuss the applicability of fuzzy pattern recognition in the context of fuzzy inference system in construction through validation results of the proposed approach in this study compared with other disciplines.

BACKGROUND

This section provides a literature review of four main relevant areas to this study: (1) an overview of decision-aid approaches in project delivery method selection; (2) the use of fuzzy sets in project delivery decision-making; (3) fuzzy pattern recognition; and (4) fuzzy inference system.

Decision-Aid Approaches in Project Delivery Method Selection

Although numerous decision-aids have been proposed to assist practitioners in project delivery method selections, these aids have traditionally been limited to either quantitative or qualitative approaches separately rather than in combination (Nguyen et al. 2020). For instance, Al Nahyan et al. (2018) proposed a decision-aid system to empirically rank delivery methods commonly used in mega projects based on qualitative variables, including project risks, project constraints, and opportunities of investments. Tran and Molenaar (2015) attempted to quantify the differences in project cost performance between project delivery methods with a risk-based approach. Qualitative approaches concentrate on identifying and investigating qualitative variables, such as levels of complexity, impacts of delivery risks, and owner's satisfaction, which are decisive factors in construction project delivery method selection (Bypaneni and Tran 2018; Choi et al. 2019). In fact, the majority of project delivery method decision-making in the construction industry is based on subjective judgements of experts and guidelines from professional organizations regarding qualitative variables (Korkmaz et al. 2010). Quantitative approaches typically rely on probabilistic statistical models to evaluate different sets of project delivery method selection criteria (Tran and Molenaar 2015) as well as empirically comparing project delivery alternatives based on historical data (Sullivan et al. 2017). However, human-related bias in using quantitative approaches to deal with qualitative inputs (e.g., ratings of risk impacts and project complexity levels) is apparent and may cause uncertainty in findings. A gap in the existing body of knowledge of project delivery method decision-making is to investigate the applicability of statistical approaches that are specifically suited to simultaneously accommodate both quantitative and qualitative variable inputs.

Fuzzy Sets in Project Delivery Decision-Making

Construction research has attempted to investigate fuzzy sets to support decision-making in project delivery selection with sophisticated and hybrid models, which may lead to difficulties in practically reproducing the research processes (Chan et al. 2009; Elbarkouky et al. 2016; Paek et al. 1992). Mostafavi and Karamouz (2010) presented a technical note discussing the use of a fuzzy approach to incorporating risk analysis in developing a project delivery selection model. However, the authors did not provide validation of the risk-based fuzzy model, which might question the applicability of this approach to the domain of project delivery selection. Khanzadi et al. (2016) developed a fuzzy Analytic Hierarchy Process (AHP) multi-criteria decision-making framework to evaluate delivery methods based on subjective judgements of a group of experts. This proposed AHP model mainly relies on weights and scores for input selection criteria. Using a dam and hydropower plant project to evaluate the applicability of the proposed model, the authors suggested using more sample projects to provide better validation of the model’s outputs (Khanzadi et al. 2016). To handle uncertainties in the project delivery selection process, Martin et al. (2017) established a forward normal cloud model based on fuzzy sets with a normal distribution membership function to represent preferences of decision makers. Validation of the fuzzy-based cloud model was confirmed by six construction professionals with an average experience of 22 years. Al Nahyan et al. (2018) proposed a fuzzy logic model in selecting an appropriate delivery method for mega projects in the United Arab Emirates (UAE) with 127 completed survey responses mainly focusing on risk assessments. Although Al Nahyan et al. (2018) employed fuzzy sets in project delivery decision-support systems by producing a decision matrix for project delivery selection, it falls short of providing an inference system to assist decision makers. In summary, the previous studies showed particular advantages of using fuzzy-based approaches to better understanding project delivery decisions, modeling uncertainty, handling imprecise measurements, accommodating combinations of different types of variables, and providing computational flexibility.

Fuzzy Pattern Recognition

Fuzzy pattern recognition has been used in many engineering fields as a promising tool to accommodate the measurement of subjective and qualitative input data (Elwood 2014). Pattern recognition is a mathematical technique to determining and understanding potential latent structures in data by comparing them to known structures, which are established through classifications (Nagalakshmi and Jyothi 2013). Different from a typical classification process, such as cluster analysis, support vector machines, and k-nearest neighbor classifier, fuzzy pattern recognition aims at systematically classifying input data to one of known data patterns (Elwood 2014). Fuzzy pattern recognition combines fuzzy sets to the recognition process to identify the most matching groups for inputs based on the mean and covariance of data features (Bezdek 1999). This fuzzy-based classification is typically used in exploratory studies because of its ability to accommodate various combinations of different types of input variables based on similarities and dissimilarities (Viattchenin et al. 2013). Commonly, data classification has been studied under probability theory with crisp sets of numbers (Hastie et al. 2009). Different from probability theory, where the frequency and likelihood of a datum in a class of number are indicated, fuzzy set theory demonstrates the similarity and dissimilarity of a datum to the class. Therefore, when classification involved uncertainties and imprecise information, the use of fuzzy sets is more appropriate than probability (Ross 2010).

Fuzzy Inference System

The algorithm of fuzzy pattern recognition is built upon a fuzzy inference system, where fuzzy set theory is used to map input data to possible output patterns (Hoppner et al. 1999). Specifically, a set of unlabeled (or unknown class) data as inputs are reasoned via a definite set of rule-based inference engines to recognize possible patterns for them as outputs. The fuzzy system is appropriate for quantifying uncertainties involving human intuitive thinking and opinions because it models ambiguity in human cognitive process and deal with imprecise information (Zeng and Starzyk 2018). One of critical elements in establishing a fuzzy inference system is to generate knowledge bases, which contain a set of fuzzy "*If-Then*" rules. Because the set of rules can be formulated by experts or extracted from historical data, fuzzy systems practically provide reasoning rules observed from empirical information. The input data sets have to be fuzzified before proceeding through any inference engines. It is noted that the use of fuzzy systems in the construction industry is mainly related to triangular and trapezoidal membership functions (Fayek and Oduba 2005).

RESEARCH MOTIVATION

Current project delivery selection predominantly depends on different combinations of quantitative and qualitative inputs (Khanzadi et al. 2016; Mafakheri et al. 2007). However, current statistical approaches face many obstacles in investigating qualitative selection criteria (Al Nahyan et al. 2018; Chen et al. 2011), which creates a gap in project delivery selection (Chan et al. 2009). Fuzzy sets in the context of fuzzy pattern recognition can support decision makers in project delivery selection by providing an appropriate means to handle different types of variables, approximate missing and imprecise information, and model uncertainty in the selection process. The fuzzy inference system can be adjusted with a relative ease by modifying either the overlapping membership functions or the fuzzy rule base, which can help with incorporating new knowledge (Elwood and Corotis 2015). By providing rule-based classifiers with the benefit of membership functions and geometric interpretability, fuzzy pattern recognition is appropriate in modeling subjective information and uncertainty in project delivery selection. For instance, the fuzzy inference system can be used as a deterministic approach to quantify qualitative independent variables (Fayek and Oduba 2005). However, no project-delivery-selection-based study has investigated the use of fuzzy pattern recognition to identify and recognize the potential common patterns of delivery methods used in highway construction. This study bridges the literature gap of project delivery selection by demonstrating and validating the application of fuzzy pattern recognition in identifying cost performance associated with two main project delivery methods used in the highway construction industry: D-B-B and D-B.

Decision-making practices for highway project delivery method selection are oftentimes subjective and have historically depended upon the experience of the agency with respect to the selection criteria that are considered. These criteria are often formed based on a combination of qualitative and quantitative variables. Common quantitative variables consist of project type, facility type, and project performance. Project type is an element factor to selecting possible delivery methods in highway projects (Tran et al. 2018) while project performance, such as cost growth, schedule growth, and quality, has been extensively used to distinguish and prioritize project delivery alternatives (Douglas et al. 2016; Sullivan et al. 2017). Common qualitative variables involve levels of project complexity and impact of delivery risks. The intuition of project complexity and ratings of delivery risk impacts greatly contribute to selection of an appropriate delivery methods in highway projects (Al Nahyan et al. 2018; Tran and Molenaar 2015). Since each construction project is unique and has a distinct set of associated quantitative and qualitative variables, the identification, collection, and investigation of common selection criteria of project delivery method selection are challenging. The use of fuzzy pattern recognition can avoid biases of subjective decisions made by the project team and accommodate different combinations of quantitative and qualitative variables.

RESEARCH QUESTIONS AND METHODOLOGY

This study attempts to investigate the following research question: How does fuzzy pattern recognition implement project delivery selection in the highway sector? To respond to this research question, the research methodology of this study includes two main phases: (1) development of a fuzzy rule-based inference system based on historical project data (i.e., training the data) and (2) verification and validation of the proposed fuzzy pattern recognition process (i.e., testing the system). Figure 1 shows an overview of the methodology. Specifically, this study first conducted a comprehensive literature review of delivery method selection and fuzzy-based applications in this domain to identify the research gaps. Then, a fuzzy rule-based inference system was developed based on six project attributes, including facility type, project type, project complexity, delivery risk, delivery method, and cost growth, in an R programming environment. Subsequently, a case example was performed to illustrate the pattern recognition capability of the system (i.e., delivery methods and cost growth) for a new highway project based on its project attributes. Finally, the study presents the results of the case example in detail and discusses the validation process.

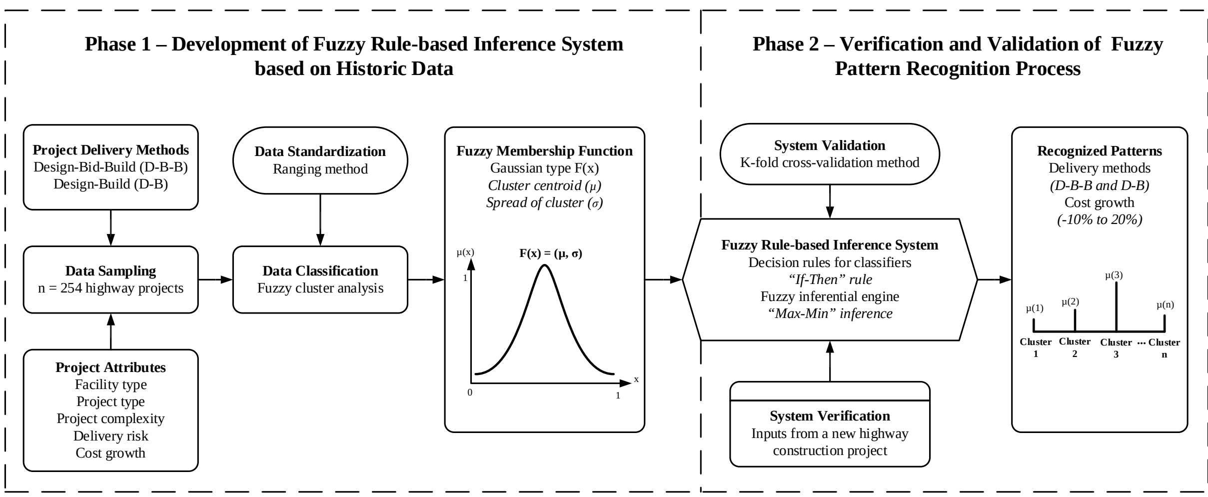

Figure 1. Research methodology

PHASE 1 – DEVELOPMENT OF FUZZY RULE-BASED INFERENCE SYSTEM

The core function of fuzzy pattern recognition refers to the use of a fuzzy inference system, which is established based on experience of experts or rules extracting from reasoning approximations. This study employed the "*If-Then*" rule-base in reasoning and recognizing a matching pattern for inputs developed by Mamdani and Assilian (1975). There are three main steps of establishing a fuzzy Mamdani inference system, including (1) fuzzification, (2) rule-based inference, and (3) defuzzification (Chiu 1997). First, fuzzy membership functions, which convert raw inputs to fuzzy values, are developed. Second, fuzzy values are aggregated with the fuzzy Mamdani rule-base inference to assign data to appropriate groups. Finally, assigned fuzzy data are converted back to crisp values. As prerequisites of the pattern recognition technique, data standardization and classification need to be executed before utilizing the main fuzzy inference system (Duda et al. 2001). To demonstrate the use of fuzzy-based techniques in this study, custom fuzzy-based programming functions were coded, developed, and employed in R programming environment based on Garibaldi et al. (2017), Knott et al. (2013), and Riza et al. (2014).

Data Sampling

This study utilized the empirical database of 291 completed highway construction projects collected from 28 state DOTs. The collected dataset contained empirical information regarding facility type, project type, delivery methods, procurement methods, payment methods, overall performance (i.e., cost, schedule, and quality), and risk profiles. Three fundamental project delivery methods including D-B-B, D-B, and CM/GC were included in the survey. However, it is noted that there was a limited number of CM/GC projects in the dataset because of CM/GC was still relatively new to many state DOTs at the time of data collection. Project cost data includes the engineer's estimate, contract award, and final cost of each collected project. Missing data and outliers critically influence the execution of fuzzy pattern recognition (Kruse et al. 2007). Therefore, boxplots and the missingness map of 291 collected highway projects were examined using R programming descriptive statistics packages in terms of project type, facility type, project complexity, delivery risks, and cost growth. Accordingly, 37 projects were identified with extreme outliers, either greater than 20% or less than -10% of cost growth. After eliminating missing data and extreme outliers, the final dataset (dataset *P* herein) included 254 projects with six project attributes: facility type, project type, project complexity, delivery risk, delivery methods, and cost performance. It is important to note there were only 27 CM/GC projects out of 254 projects (10.6%). As a result, CM/GC is not a dominant delivery method in this sample. More discussion on project delivery patterns can be found at Nguyen et al. (2020).

The facility type variable was measured by estimated percentages of the total project cost ranging from 0 to 100% and consisted of five sub-categories: road, bridge, drainage, intelligent transportation system (ITS), and other facility types. ITS is an advanced system which aims to improve traffic efficiency, safety, and mobility by providing innovative transportation services, such as sensing and control technologies, local real-time traffic information, and transportation network analysis (Zhankaziev et al. 2018). The other facility types include temporary traffic controls, beacons, pedestrian trails, park-and-ride, retaining walls, tolling structures, and bridge coatings. The project type was measured using three sub-categories: new construction, reconstruction, and other project types. The other project types include guardrail repair safety maintenance, bridge replacement, stream restoration, road widening, and new sidewalk. The level of project complexity was measured based on a 3-point ordinal scale and defined by “most complex”, “moderately complex”, and “non-complex”. The “most complex” group includes major scopes of work, such as new highways, new interchanges, capacity adding/major widening, and large reconstruction projects, and requires congestion management studies as well as the statement of environmental impacts. The “moderately complex” group includes reconstruction projects which do not add capacity, such as minor roadway relocations, non-complex bridge replacements, and requires minor environment assessment or categorical exclusion. The “non-complex” group includes maintenance betterment, overlay projects, simple widening without right-of-way and little or no utility coordination, non-complex enhancement projects without new bridges, and categorical exclusion.

The risk profile consists of thirty-one project risks rated with a 5-point ordinal scale, which represents five potential degrees of risk impacts (i.e., very low, low, moderate, high, and very high) on cost performance. This study utilized the result of exploratory factor analysis (EFA) from Bypaneni (2017) and Mathew et al. (2018) to reduce 31 risk factors to seven underlying risk factors which are commonly impacting the selection of project delivery methods in highway construction. Bypaneni (2017) found the factor loadings of seven critical risk factors within D-B-B and D-B using EFA and established the interrelationships between them to support selection of project delivery methods. Mathew et al. (2018) attempted to identify the relationship between delivery risks and cost growth in highway construction projects using structural equation modeling. The resulting seven risk factors used in this study were defined as follows:

1) Complexity risk factor contains project complexity, uncertainty in geotechnical investigation, legal challenges, and changes in law, intergovernmental agreements and jurisdiction, and difficulty in obtaining other agencies.
2) Quality risk factor contains construction quality control and quality assessment process and design quality assurance.
3) Constructability risk factor contains delays in procuring critical materials, labor, and specialized equipment and significant increase in material, labor and equipment cost.
4) Construction risk factor contains work zone traffic control and construction sequencing, staging, and phasing.

5) Utility and right-of-way (ROW) risk factor contains unexpected utility encounter, delays in completing utility agreements, and delays in ROW process.
6) Management risk factor contains staff experience and availability, project and program management issues, and conformance with regulations, guidelines, and design criteria.
7) Environmental risk factor contains challenges to obtain appropriate environmental documentation and environmental impacts.

The cost performance variable was considered in terms of cost growth at the project completion. It is the percentage change in cost between the contract awarded amount and the final cost. The final cost was the sum of the contract award value and costs of all change orders while the contract award is the total amount indicated in the signed contract.

To develop the fuzzy pattern recognition, six project attributes were used, including facility type, project type, project complexity, delivery risk, and project cost growth. Each project attribute has a particular number of sub-categories as discussed above. Project delivery methods (e.g., D-B-B and D-B) were used as control variables. As a result, a set of 17 variables were used for developing fuzzy pattern recognition described as follows:

- Facility type (m_1 = 5): road, bridge, drainage, ITS, and others.
- Project type (m_2 = 3): new construction, reconstruction, others.
- Project complexity (m_3 = 1): most complex, moderately complex, and non-complex highway projects.
- Delivery risk (m_4 = 7): complexity risk factor, quality risk factor, constructability risk factor, construction risk factor, utility and ROW risk factor, management risk factor, and environmental risk factor.
- Cost performance (m_5 = 1): cost growth.

It is noted that the number of selected features practically affects the accuracy of the classification and pattern recognition results (Elwood 2014). This study focused on demonstrating the practical use of this fuzzy-based technique in early project delivery decision-making in highway construction. Therefore, the selected variables include common project attributes (i.e., facility type, project type, project complexity, delivery methods, and risk factors) and project cost performance. This selection is supported by the literature. For example, Al Nahyan et al. (2018) stated that project type and complexity are typically recorded in aiding selection of appropriate delivery methods while Creedy et al. (2010) and Mostafavi and Karamouz (2010) indicated risk factors play an important role in the pool of project delivery selection criteria. Khanzadi et al. (2016) and Touran et al. (2011) recommend the use of cost performance in any project delivery decision-making processes.

Data Standardization and Classification

To guarantee the validity of classification-based applications, raw data need to be standardized to obtain the same measuring scale (Ross 2010). In this study, raw data were standardized by using Min-Max normalization, a method produces a unified numerical scale (i.e., 0 to 1) for raw values by ranging (Elwood and Corotis 2015). Kassambara (2017) recommends the use of ranging approach with the min and max values of each variable to preserve the relationships among data by treating separately based on its own range and sample. Accordingly, normalized data of 17 selected variables were ranged from 0 to 1 and proceeded with fuzzy cluster analysis to identify typical groups of highway projects in terms of delivery methods and cost growth. The qualitative variables in this study, including project complexity and delivery risk factors, were normalized before being assigned to clusters representing particular attributes. Based on the distances between projects (i.e., data points) within clusters, membership values of each project to specific variables (i.e., qualitative – project complexity and delivery risk factors, and quantitative – project type, facility type, cost growth) were given.

Prior to implementing fuzzy pattern recognition to the collected dataset, typical patterns of cost performance were identified by using fuzzy classification in the context of fuzzy cluster analysis. This technique helps allocate common groupings of selected variables (Novák et al. 2016). The focus of this study was to develop a fuzzy rule-based inference system from clustering results. Membership functions and "*If-Then*" rules are formulated directly from the data clusters with the cluster centers performing as representations for fuzzy rules. The analysis provides a powerful mean to investigate and assess the fuzzy rules in multiple dimensions (Castellano et al. 2007). Fuzzy cluster analysis not only generates the rule base but also designs classifiers, which represent patterns for recognizing input data. The theoretical procedure of using fuzzy cluster analysis can be found in Elwood (2014) and Ross (2010).

Table 1 shows seven groups classified based on the degree of similarity of project attributes, delivery methods, and cost performance. Each fuzzy cluster represents a typical set of highly similar project attributes, selected delivery methods associated with cost growth, based on the cluster center and the membership degree of data points within the clusters. For example, cluster 2 (D-B_{Low}), contains D-B highway projects with low cost growth. The membership degree of each project to a specific cluster was measured by the distance between the project and the center of that cluster. The number of clusters was subjectively determined based on the goodness of clustering results via cluster validity indices (Lantz 2015). Using five typical fuzzy cluster validity indices shown in Das (2013), seven clusters demonstrated the most reasonable groupings within the collected dataset. Mathematically, the procedure of fuzzy pattern recognition proposed in this study was developed based upon the collected dataset P consisting of 254 highway projects, $P = \{P_1, P_2, P_3, ..., P_{254}\}$. Each project P_i has 17 variables (e.g., five sub-facility types, three sub-project types, project complexity, seven risk factors, and cost growth) $P_i = \{P_{i1}, P_{i2}, P_{i3}, P_{i4}, ..., P_{i17}\}$. A set of seven predefined clusters serve as classifiers in the fuzzy rule-based inference system $C_j = \{1, 2, 3, ..., 7\}$.

Table 1. Predefined clusters of selected project attributes

No.	Cluster	N	Facility Type	Project Type	Project Complexity	Risk Impact
1	D-B-B$_{\text{Saving-Low}}$	52	Road	Reconstruction	Moderate	*VL to L
2	D-B$_{\text{Low}}$	34	Bridge	New construction	Moderate	VL to L
3	D-B-B$_{\text{Medium-High}}$	39	Road	New construction	Most	VL
4	D-B$_{\text{Saving-Low}}$	34	Road	New construction	Most	L to M
5	D-B$_{\text{Low}}$	23	Road	Combination	Moderate	VL to L
6	D-B-B$_{\text{None}}$ and D-B$_{\text{None}}$	39	Bridge	Reconstruction	Moderate	VL to L
7	D-B$_{\text{None-Low}}$	33	Road and Bridge	New construction	Most	VL to L

**VL – Very Low; L – Low; M – Medium*

Fuzzification

The fuzzifying process converts input values into standardized fuzzy numbers prior to the inferential reasoning process. In this study, each input project is represented as a multi-dimensional data vector, consisting of the number of selected variables. The objective of the fuzzification process is to manipulate input data with Gaussian fuzzy membership function to convert raw inputs and insert membership values (Hwang 2004). The Gaussian-based functions are more appropriate than other types of membership functions, including triangular, trapezoidal, and sigmoid to handle a combination of quantitative and qualitative inputs (Daneshvar 2011). The Gaussian-based membership function has the maximum value of 1. The linguistic labels were assigned to each approximated membership function of 17 selected variables based on the values of cluster centers.

Rule-based Inference System

A rule base is projected from the membership function and written in terms of the input attributes and output group labels (i.e., rule antecedents and rule consequents, respectively) (Setnes et al. 1998). In the fuzzy rule-based system, each rule demonstrates a membership value that exhibits the extent to which the project attributes belong to a cluster C_i. With pre-defined groups partitioned with fuzzy cluster analysis (i.e., input values were ready in terms of fuzzy sets), the Mamdani "*If-Then*" fuzzy inference system is recommended to use for identifying relationships between the cluster center and within-cluster projects (Mamdani and Assilian 1975). For example, if the membership value of project P_1 shown in Cluster C_1 is higher than the membership values of this project in other clusters within a 17-D space then project P_1 belongs to Cluster C_1. In other words, in the "Road" variable space, the highest membership value of project P_1 is identified in Cluster C_1; similar to the other 16 variable spaces (e.g., bridge, drainage, ITS, risk factors 1 to 7). Another example, if project P_2 has its highest membership values in terms of 17 variables in cluster C_2 compared to other clusters, project P_2 belongs to cluster C_2. Given a new highway project with membership values of particular variables in each cluster, the inference system identifies the cluster that contains the given project based on a pre-defined set of rules. That is, a rule points out the cluster that shows the highest membership values of that project in the variable spaces. To maintain the interpretability of fuzzy inference systems in the domain of pattern recognition, the number of input attributes and the number of fuzzy rules should be limited to 7 ± 2 (Castellano et al. 2007). The "*If-Then*" rule aims to match the input variables to the most likely pattern with the highest membership values.

Defuzzification

Fuzzy values obtained from the rule-based inferential system are converted to crisp values via the defuzzification process. The rule-based classifier represents a nonlinear relationship between the inputs (i.e., project attributes) and outputs (i.e., clustering groups). After recognizing the matching pattern for the input project, fuzzy results were converted to be crisp values using the variation of the proposed ranging method to maximize the interpretability of the fuzzy rule-based inference system.

PHASE 2 – VERIFICATION AND VALIDATION OF FUZZY PATTERN RECOGNITION

This phase consists of two main tasks. First, the verification of the proposed fuzzy inference system was presented using a case example in which possible patterns of delivery methods and relevant cost growth were recognized for a randomly selected highway project. Second, the validation of the proposed system was performed with k-fold cross-validation. This section aims at providing a step-by-step guideline to help practitioners select an appropriate delivery method by using the proposed fuzzy system for any given set of project input attributes.

Fuzzy System Verification

To demonstrate the step-by-step process of the fuzzy rule-based inference system, the authors randomly selected a highway project in dataset *P*: a new construction road project, "SR25 Hoosier Heartland Highway" (Case Project hereafter), in Indiana. This project was characterized as a four-lane divided limited-access rural highway. The total project cost was $36,424,873. There were 38 issued change-orders related to the final quality adjustment resulting in an increase of $1,235,701. The main reason for a total cost growth of 18% was that this project faced a high level of cost impact from complexity- and quality-related risks.

Step 1 – Fuzzifying Inputs

Based on the data collected, Table 2 summarizes Case Project's attributes, including facility type, project type, project complexity, and delivery risk factors, as raw and standardized inputs.

Table 2. Example of project inputs to Fuzzification process

No.	Variables	Range	Raw Inputs	Standardized Inputs
1	Road	0 - 100%	80%	0.80
2	Bridge	0 - 100%	20%	0.20
3	Drainage	0 - 100%	0%	0
4	ITS	0 - 100%	0%	0
5	Other Facility Types	0 - 100%	0%	0
6	New Construction	0 - 100%	100%	0.10
7	Reconstruction	0 - 100%	0%	0
8	Other Project Types	0 - 100%	0%	0
9	Project Complexity	1 - 3	1	0
10	RF1 - Complexity Risk Factor	1 - 6	2	0.20
11	RF2 - Quality Risk Factor	1 - 6	2	0.20
12	RF3 - Constructability Risk Factor	1 - 6	2	0.20
13	RF4 - Construction Risk Factor	1 - 6	2	0.20
14	RF5 - Utility and ROW Risk Factor	1 - 6	2	0.20
15	RF6 - Management Risk Factor	1 - 6	2	0.20
16	RF7 - Environmental Risk Factor	1 - 6	1	0.10

Gaussian membership function was used to fuzzify input variables, consisting of facility type, project type, project complexity, and risk factors (Table 2). Project attributes of a new highway project were entered to the fuzzy system by using a developed graphical user interface (GUI) with predefined attributes along with associated Gaussian membership functions. For instance, Case Project's attributes were entered using GUI, including facility type (road = 80%, bridge = 20%), project type (new construction = 100%), project complexity (most), risk factor 1 (very low), risk factor 2 (very low), risk factor 3 (very low), risk factor 4 (low), risk factor 5 (low), risk factor 6 (very low), and risk factor 7 (low). Figure 2 displays project inputs to the fuzzy inference system. Afterwards, the standardized inputs were proceeded using the Mamdani rule-based inference system.

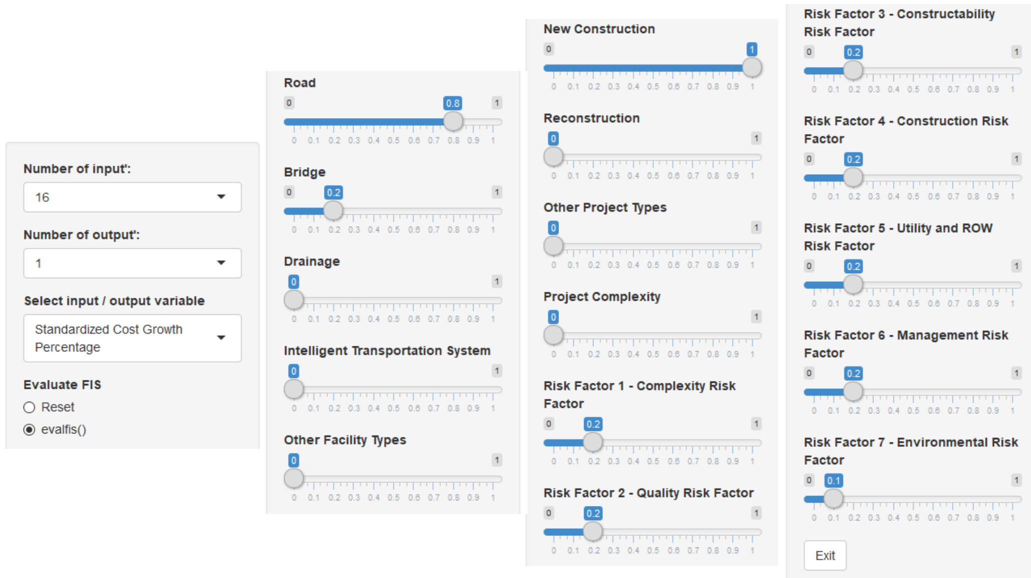

Figure 2. Example of project inputs to fuzzy inference system using GUI

There were two primary inputs for the algorithm of fuzzy pattern recognition. The first input was the matrix of membership values that assigns relevant membership values to each defined cluster. The second input was a vector of cluster centers that provides a geometric relationship with individual membership values which are estimated with fuzzy membership functions. Equation (1) shows an example of the center value of cluster 1 with the selected variable “Road”. This formula is essentially replicated for each variable to define the overall value of the cluster. In other words, the membership value of Case Project in cluster 1 in the “Road” space was summed up with the other 253 membership values to calculate the center value of cluster 1 in the “Road” space. The overall combination (i.e., the vector of cluster centers) was established in a 17-dimensional space (i.e., a space built upon 17 selected project variables) per cluster to produce the outputs.

$$Cluster\_center_{1-Road} = \frac{\sum_{i=1}^{254}\left[\left(membership\_value_{1i-Road}\right)^{17} \times P_{i-Road}(\%)\right]}{\sum_{i=1}^{254}\left(membership\_value_{1i-Road}\right)^{17}} \quad (1)$$

In this study, the vector of cluster centers contained seven cluster centers developed based on a 17-D space while the matrix of membership values illustrated the degree of membership of each project to each selected attribute within a particular group of cost performance. The inputs of 17 selected project variables are established by approximating the rows of the matrix of membership values; for example, $\mu_{1-Road}\left(P_{Case}\right) = \mu_{1-Case}$ with $\mu_{1-Road}\left(P_{Case}\right)$, the degree to which Case Project belongs to cluster 1 based on the project variable "Road", as shown in Eq. (2):

$$\mu_{1-Road}\left(P_{Case}\right) = \exp\left[-\frac{1}{2}\left(\frac{P_{Case-Road} - \mu_{1-Road}}{\sigma_{1-Road}}\right)\right] \quad (2)$$

Where: μ_{1-Road} is the centroid value of cluster 1 in the "Road" variable space (i.e., the maximum value of all memberships in cluster 1) and σ_{1-Road} is the spread of the membership function (i.e., fuzziness of the cluster), which demonstrates the distance from the center of cluster 1 to the farthest project assigned to cluster 1. This equation is replicated for all seven clusters in terms of 17 selected project variables. Figure 3 shows an example of the Gaussian membership function for the "Road" variable within cluster 1. Each project that has the feature of "Road" involves in cluster 1 with a particular degree of membership. On the horizontal axis, the Gaussian membership function describes the standardized value of each project regarding the road's percentage of it. On the vertical axis, the membership value of each project that belongs to cluster 1 with the representative road-type is provided on a scale of 0 to 1.

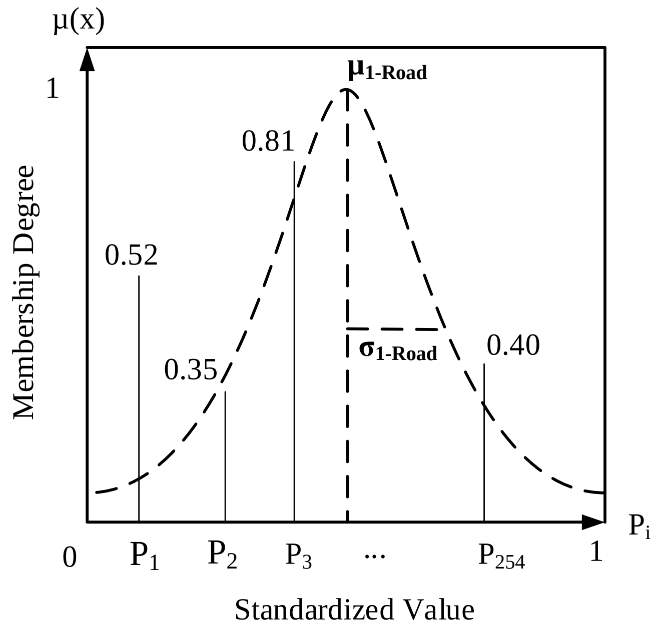

Figure 3. Example of Gaussian membership function of "Road" variable in cluster 1

Step 2 – Recognizing Patterns

After defining the inputs (i.e., attributes of Case Project) to the fuzzy inference interface, the project attributes were assigned to seven clusters based on the degree of membership values to the associated membership function of each cluster. The Mamdani "*If-Then*" rule base implemented in this study is strictly adhere to the Min-Max inference method, an aggregating approach proposed by Mamdani and Assilian (1975) for fuzzy-based inference systems. In Min-Max inference, the "*and*" conjunction in each rule is evaluated by the fuzzy set minimum operator and the aggregation over all the rules is evaluated by the maximum operator (Elwood and Corotis 2015). Accordingly, this method, first, calculated the membership values of seventeen variables, and then selected the minimum value of each variable within seven clusters. Based on the identified minimum values, the recognized pattern was the one that obtains the maximum values of degree of membership among seven clusters. This inference assesses if the input project satisfies the minimum requirement of the recognized cluster along with obtaining the maximum degree of belongingness of the project to that cluster. The degree of membership of Case Project in the "Road" variable presented in that variable's projection space is $\mu_{1-Road}(P_{Case}) \in [0,1]$. The membership values in each cluster that describes the degree of similarity are denoted by $\mu_{Ci}\left(P_1, P_2, P_3, ..., P_{254}\right) = \min\left\{\mu_{1-Road}(P_i)\right\}$ for i = 1 to 254. Aggregation over all of the individual rules was denoted by $\mu_C\left(P_1, P_2, P_3, ..., P_{254}\right) = \max\left\{\mu_{C1}, \mu_{C2}, \mu_{C3}, ..., \mu_{C7} \mid C_i = C\right\}$. Where μ_C is the membership degree to which the vector of Case Project's variables ($P_{Case-Road}$, $P_{Case-Bridge}$, $P_{Case-Drainage}$, …, $P_{Case-RF5}$, $P_{Case-RF6}$, and $P_{Case-RF7}$) is assigned to a cluster C_1. Figure 4 illustrates a detailed process of implementing the fuzzy rule-based inference system with five main predominant attributes and seven predefined clusters.

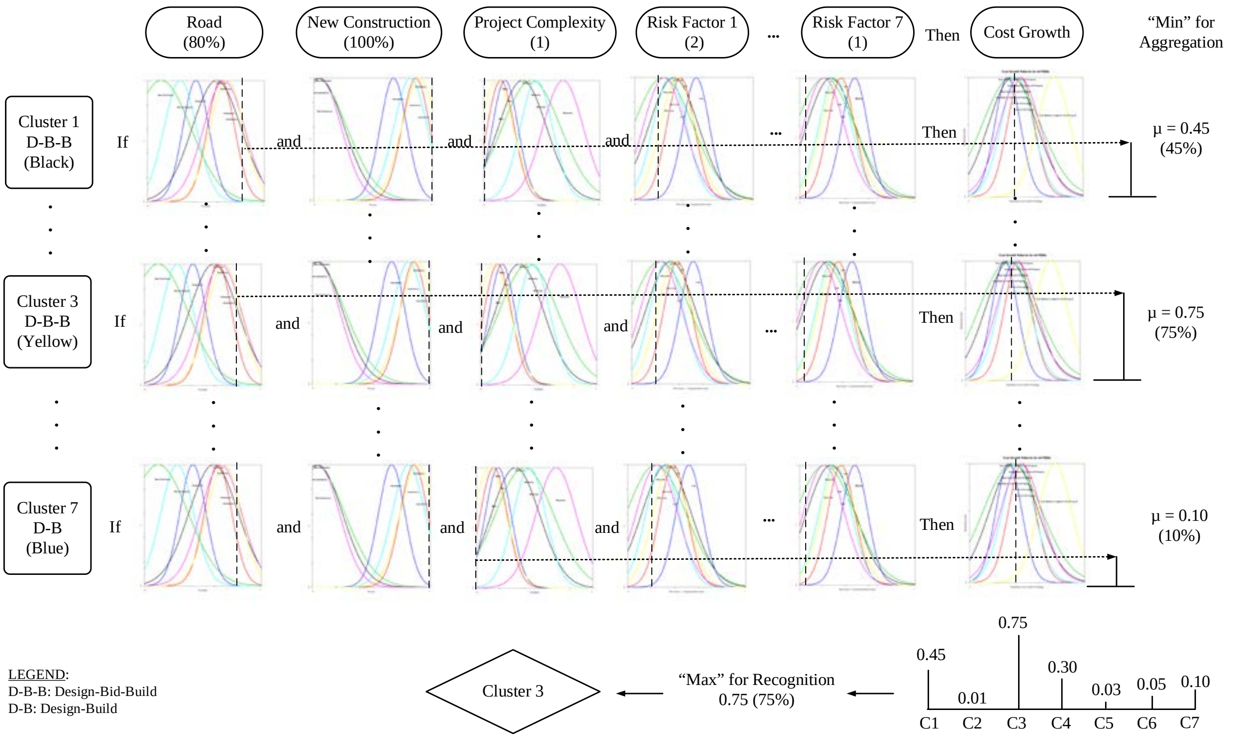

Figure 4. Illustration of fuzzy pattern recognition process

Step 3 – Defuzzzifying Outputs

The fuzzy membership values were defuzzified before interpreting the outputs of recognized patterns. This task must be taken place prior to computing and evaluating classification error rates. The defuzzifying process generates a recognized pattern for input projects using Eq. (3) (Emily and Corotis 2015). This equation provides the mechanism of assigning a vector of input project attributes (P_{1-Road} , $P_{1-Bridge}$, $P_{1-Drainage}$, …, P_{1-RF7}) to the clustering group where the project represented the highest membership.

$$C\left(P_1, P_2, P_3, \ldots, P_{254}\right) = \mathrm{C_i} \text{ if } \mu_{C_i}(P_1, P_2, P_3, \ldots, P_{254}) > \mu_{C_j}(P_1, P_2, P_3, \ldots, P_{254}) \text{ for all } \mathrm{C_i} \neq \mathrm{C_j} \tag{3}$$

For Case Project, Eq. (3) was written as: $C\left(P_{Case}\right) = \mathrm{C_3}$ if $\mu_{C_3}(P_{Case}) > \mu_{C_j}(P_{Case})$ for all $\mathrm{C_3} \neq \mathrm{C_j}$. Figure 4 shows that Case Project represented the highest membership values within cluster

3 (75% of degree of membership) compared to other clusters. Therefore, Case Project belonged to cluster 3 where information about the typical delivery method used associated with cost growth was provided. Using the ranging conversion method, standardized values of 17 variables of this project were converted to crisp values for the interpretation purposes. As a result, Case Project, a new construction road project with the highest level of project complexity, was delivered by D-B-B and resulted in a medium to high cost growth (10% - 20%). Through comparing with the actual project delivery method used and project performance, the result indicates that the fuzzy classifiers recognized the correct grouping for this project.

Fuzzy System Validation

This study utilized a k-fold cross-validation method to validate the fuzzy system where the collected data were iteratively computed with the matching rates of the recognizing patterns. The k-fold validation method is commonly used in the domain of fuzzy pattern recognition (Piegat 2001). Conducting this method, the collected data were divided into two main sets: training and testing. The training set (i.e. consisting of 90% of the entire dataset) was used to formulate parameters of the fuzzy classifiers (i.e., design the fuzzy inference system). The testing set (i.e., consisting of 10% of the entire dataset) was used to assess the rate of misclassification and compute the degree of recognizing errors (i.e., validate the capability of the fuzzy system). The training and testing error rates are computed by counting the number of misclassifications of data in the fuzzy classifiers based on the degree of membership (Elwood 2014). The training error rate ($E_{training}$) is used to assess if the model validity and computed by finding the ration of $N_{incorrect}$ to $N_{training}$, where $N_{incorrect}$ is the number of misclassified projects; $N_{training}$ is the total number of projects in the training set. The testing error rate ($E_{testing}$) is used to assess if the model reliability and computed by finding the ratio of $N_{incorrect}$ to $N_{testing}$, where $N_{incorrect}$ is the number of misclassified projects and $N_{testing}$ is the total number of projects in the testing set.

Because a 10-fold cross-validation was used in this study, the ratio of the training set to the testing set should follow a 90:10 split (Hastie et al. 2009). The 90:10 split is used in validating the applications of fuzzy pattern recognition in other fields, such as mechanical engineering (Tran et al. 2009) and computer engineering (Ye et al. 2006). Accordingly, the training set consists of 228 projects (i.e., 90% of the entire dataset) while the testing set contains 26 projects (i.e., 10% of the entire dataset). Specifically, 10 different combinations of each set were run and replaced iteratively. The variation between 10 separate runs was low (less than 2%). This study utilized fuzzy cluster analysis with fuzzy C-Means algorithm prior to the fuzzy pattern recognition process, so the sensitivity to noise of the developed inference system becomes less (Ross 2010). After running 10 iterations of the validation and testing sets, the average training and testing error rates of the fuzzy inference system were recorded. From the training set, 31.3% of the clustered projects suffered low memberships compared to the actual projects. From the testing set, 29.6% of the clustered projects suffered low memberships compared to the actual projects. It is noticed that this result can in part be explained by the definition of fuzzy set theory and pattern recognition, where each data point has a partial degree of membership to the recognized cluster (Bezdek 1999). In addition, because the clustering results are typically extracted from an *n*-dimensional space (i.e., 17 dimensions in this study), no data point has a full membership to a specific cluster (Niskanen 2004; Ross 2010). As a result, the interpretation of the fuzzy recognizing results should be referred to the highest value of data points in the identified cluster.

DISCUSSION

This study demonstrated that fuzzy inference systems can serve as a decision-aid in selecting an appropriate delivery method in highway design and construction projects. This approach attempts to systematically map the inputs of a new highway project to one or more fuzzy classifiers based on the degree of similarity in project variables and then provide information about delivery methods used associated with cost growth. For example, if a new highway project has a similar set of project variables to Case Project, the system recognized the pattern of D-B-B with a likelihood of 75% to produce an average cost growth of 15%. Based on the recognition result, adjustments of the inputs can be made based on decision maker's judgements on the level of risk impacts and project complexity, to come up with a more appropriate delivery method. For instance, the impacts of seven risk factors used in the inference process can be adjusted by sliding up and down the inputs from the "risk factor" bars in GUI. Different judgements on seven risk factors may lead to multiple decision scenarios with different recognized patterns of delivery methods associated with cost growth to aid the decision maker to select a suitable delivery method. The Case Example illustrates a simple and easy-to-use procedure with predefined historic patterns of delivery methods (D-B-B and D-B) associated with cost growth performance.

The proposed fuzzy system aids decision makers in accommodating and adjusting subjective inputs in project delivery selection (i.e., project complexity and risk impact) to meet their needs. In many cases, highway projects with similar facility types and project types may end up with different cost growths and delivery methods used due to differences in other qualitative factors, such as project complexity and delivery risk. For instance, considering Case Project in two different scenarios of inputting data as shown in Figure 5: (5a) actual project attributes and (5b) similar project attributes but higher risk impacts. Raising the level of risk impacts as new inputs to the system leads to a new recognized group of cost growth and delivery method. With new inputs from seven risk factors, Cluster 4 was recognized with the highest membership value of 82%. The classified cost growths of two scenarios were exclusively different because of different degree of risk impacts. In scenario 5a, the membership values of cost growth were unstable and produced a “bimodal” distribution where the first peak was at A, B, and C and the second peak was at D, E, and F (Figure 5a). On the other hand, scenario 5b provided stable membership values of cost growth with a “normal” distribution where B, C, and D were merged. Specifically, after converting to crisp values, scenario 5b had a lower cost growth average (2%) than scenario 5a (18%). In addition, the cost growth range in scenario 5a also had a larger variation (15%) than scenario 5b (5%). Different from scenario 5a, where Cluster 3 was matched to Case Project, the system in scenario 5b recognized Cluster 4, which contains highway projects with low cost growth and delivered by D-B. These results show that this system is able to consider the adjustment of subjective inputs based on experience and preferences of the decision makers (e.g., their risk attitudes and confidence of project complexity) to recommend an appropriate project delivery method for a new highway project.

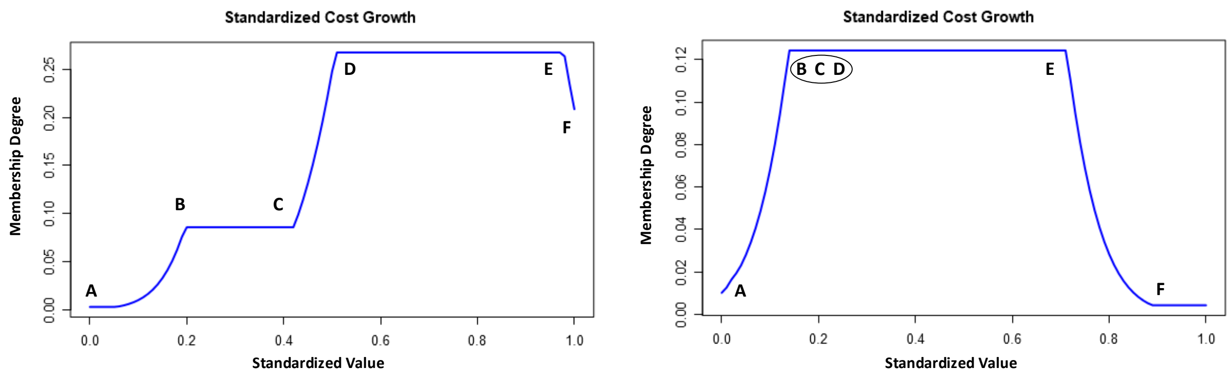

Figure 5. Two output scenarios of recognized patterns:
(5a) Actual inputs resulting in "bimodal" distribution.
(5b) Adjusted inputs with higher risk impacts resulting in "normal" distribution.

In current project delivery decision-making, both quantitative and qualitative approaches have been used to establish ranking models, compare project performance, and demonstrate simulations to determine the most optimal delivery methods (Al Nahyan et al. 2018; Chen et al. 2011; Mafakheri et al. 2007; Shrestha et al. 2012; Sullivan et al. 2017; Touran et al. 2011; Tran and Molenaar 2015). The proposed fuzzy pattern recognition approach in this study does not attempt to either rank or compare project delivery methods based on a set of selection criteria. Rather, it systematically maps the empirical patterns of delivery methods associated with cost growth to new highway projects based on similarities in project variables to recommend an appropriate delivery method based on empirical data. This approach also does not quantify any selection criteria, but relationships between chosen project variables, both quantitative and qualitative, were modeled in terms of fuzzy membership functions to explore the degree of similarities between them. Because of the inherent overlaps of project variables between fuzzy classifiers, a new project can belong to more than one cluster. Therefore, the representation of each fuzzy classifier relies on the control variable, type of delivery methods (D-B-B and D-B) and the output of cost growth's ranges and means. Among the 17 input variables, the seven risk factors produce a high sensitivity to the recognized patterns of project delivery methods and cost growth. The Gaussian membership functions associated with the seven risk factors have the highest degree of overlaps, which leads to elasticity in changing between clusters and output attributes based upon the input project's risk profile (Iqbal et al. 2015). For example, given the project attributes of Cluster 7 (new road projects delivered by D-B) as inputs with changing the seven risk factors' ratings to 2 instead of 1, the system produces the output pattern of Cluster 3 (new road projects delivered by D-B-B).

It is important to note several reasons for the error rates. First, a large number of input features and the number of clusters may create confusion and ambiguity in applications of fuzzy pattern recognition (Elwood 2014). Fuzzy-based research emphasizes that the optimal selection of the number of clusters and input variables is application-dependent (Bezdek et al. 1999). This study selected a total number of 17 features that might cause counterproductive problems to the fuzzy rule-based models. In fact, the number of selected features and the number of identified clusters should each be limited to 7 ± 2 (Castellano et al. 2007; Zeng and Starzyk 2018). Therefore, future applications of fuzzy pattern recognition in delivery-selection-based studies should consider this restriction in selecting appropriate features as well as in clustering datasets. Second, bias in collected data, which is normal in any decision-making frameworks because of uncertainty in human-related activities. Clustering assignment of a data point also depends heavily on the clustering assignment of other data points in the same cluster. Thus, appearance of biased data points can critically affect the classification leading to high error rates in the subsequent pattern recognition results. However, in the context of fuzzy set theory, it is acceptable because fuzzy sets are non-interactive, and there is restriction of a membership value based on assignment of other membership values. For instance, in structural engineering, a fuzzy-based system to recognize seismic damage for concrete structures developed by Elwood and Corotis (2015) using 136 projects and testing with confusion matrices observed the training error and testing error rates of 27% and 78%, respectively. Similarly testing with k-fold cross-validation in mechanical engineering, a fuzzy inference system developed by Tran et al. (2009) with 90 motor loads resulted in a testing error rate of 23.3%. Further, Ross (2010) concluded that the validation of fuzzy-based applications is not assessed in the context of numerical error measures as long as the computed results are rational. These relevant fuzzy-based studies in other fields support the validation results of this study that emphasizes implications for construction practices.

CONCLUSION

This study investigated the use of fuzzy pattern recognition in identifying an appropriate project delivery method for new highway construction projects. The developed graphical user interface provides an example of how fuzzy pattern recognition can be used by practitioners to accommodate a combination of quantitative (e.g., facility type and project type) and qualitative (level of complexity and risk impacts) variables in project delivery selection. Based on seven predefined groups of highway projects, a fuzzy inference system was developed using 17 variables and seven "If-Then" rules to produce empirical patterns of delivery methods (D-B-B and D-B) associated with cost growth. A 90:10 split ratio was used to train (228 highway projects) and test (26 highway projects). To further verify the applicability of the developed system, a case example was conducted with a randomly selected highway construction project in the testing set. The result shows the correct cluster of the selected project with a likelihood of

75% to produce an average cost growth of 15%. In addition, a 10-fold cross-validation technique was utilized to examine the reliability of the proposed fuzzy system. The resulting variation between the 10 separate runs was low (less than 2%). Accordingly, 29.6% of the clustered projects suffered low memberships compared to the actual projects. This validation result is well in line with other fuzzy-related studies in other fields, which supports the purpose of this study to demonstrate the use of fuzzy pattern recognition in project delivery method selection.

The proposed fuzzy rule-based inference system accounts for uncertainty and imprecision in the decision-making by standardizing all input variables and interpreting them in the context of linguistic expressions. In addition to standardization, the fuzzy inference system is also proceeded with a uniform numerical scale that eliminates the impact of data and modeling uncertainty (Elwood 2014). Although the system can only recognize patterns of the input project based on seven predefined clusters, it can explicitly measure the degree of appropriateness of the project to each cluster. Modification of the fuzzy rules are also straightforward according to the decision maker’s preference and experience. In addition, the proposed fuzzy rule-based inference system provides sufficient information about degrees of membership in each classified group and each feature, which help measure suitability of each delivery method to specific project attributes. It is important to note that although this fuzzy-based technique produces a precise level of detail in addressing uncertainty in project delivery decision-making problems, it still suffers a bias of human-involved processes. There is a trade-off between ability to accommodate various combinations of different types of variables and degrees of overlapping in the recognizing results within the domain of fuzzy pattern recognition.

Research Contributions

This study contributes to the body of knowledge by developing a fuzzy inference system to recognize possible groups of delivery methods associated with cost growth for new highway projects. It is one of the first attempts to apply fuzzy pattern recognition in the project delivery literature to leverage the evaluation of a combination of quantitative and qualitative selection criteria, which makes this study unique compared with traditional decision-aid approaches. Specifically, a mathematical approach to identifying the possible pattern of new highway projects based on their attributes and selected delivery methods was proposed. The flexibility in controlling input variables of this framework can help decision makers define and adjust critical project attributes and selection criteria. For example, the highway agencies are able to add important variables based on their preferences or remove unnecessary variables to develop a customized fuzzy inference system to recognize cost performance patterns for new highway construction projects. In addition, this data-driven fuzzy inference system also allows for multiple decision scenarios based on the decision maker's judgements of delivery risks and project complexity.

To practitioners, the proposed approach helps determine the most appropriate delivery method for a new highway project based upon project attributes and cost performance with available historical data. This study provides a decision-making tool to enhance subjective probability in the project delivery selection process. The proposed framework also helps the project team evaluate the degree of appropriateness between a new highway project with available project delivery options. By replicating the process of fuzzy pattern recognition presented in this study, industry practitioners can investigate their own empirical datasets to recognize possible patterns of data in supporting their project delivery decision.

Limitations and Future Work

There are several limitations in this study. First, the number of classifiers (i.e., the historic groups of cost growth based on delivery methods) was limited due to the available highway project data. More project data may need to be collected in future research to enhance the reliability in supporting project delivery selection. In addition, this study applied fuzzy pattern recognition to two common project delivery methods D-B-B and D-B because of the limited amount of CM/GC highway project data available. Future research may need to collect more completed CM/GC highway projects to overcome this limitation. Second, this study only considered highway construction projects to develop the fuzzy inference system. Other vertical projects and infrastructure projects can also be used to perform a similar analysis. In addition, other project attributes, such as project size, procurement method, and payment method can be included to enhance the inference system. Finally, this study only investigated project cost performance (e.g., cost growth) for the project delivery fuzzy pattern recognition. Other dependent variables such as schedule performance and quality can be incorporated in the inference system to enhance the delivery decision. This study removed 37 projects which were identified as outliers because fuzzy-based methods are very sensitive to outliers (Kruse et al. 2007; Ross 2010). Future work may analyze the reduced model prediction due to outlier removal; however, such analysis was beyond the scope of the current study. Future research may also concentrate on leveraging the developed fuzzy inference system to extend the application of fuzzy pattern recognition in the construction industry.

CHAPTER 4:

FUZZY BAYESIAN NETWORKS

INTRODUCTION

Design-bid-build (D-B-B) and design-build (D-B) have been used to deliver highway construction projects for many years. Each has advantages and disadvantages that directly impact project performance. In addition, each delivery method is attached with a number of risks and level of complexity that are subjectively assessed by experts. Selection of project delivery methods is usually conducted in the early stages of a project with the involvement of uncertainty and imprecise information that makes this decision-making process become more complex. In delivery method selection, construction experts are likely to provide a possible range of numerical values, a linguistic expression or subjective judgement (e.g., true, false, high, medium, and low) of particular selection criteria, or a fuzzy number of probabilistic uncertainties. To comprehensively select an appropriate project delivery method for new highway projects, there is a need of scientific approach to analyzing logical relationships between project attributes and updating with posterior probability for subjective variables (e.g., project complexity and delivery risks).

Bayesian networks, a probabilistic graphical model for graphically representing the relationships among a set of variables, are capable of computing the probabilities of a project attribute event under given evidence (Bayraktar and Hastak 2009; Khodakarami and Abdi 2014; Luu et al. 2009; Thanathornwong 2018; Yang et al. 2006). However, Bayesian applications in selection of project delivery methods are limited because of the prerequisite requirement of precise input information (e.g., prior probability and causal relationships of node variables) (Bakht and El-Diraby 2015; Hastie et al. 2009). In fact, it is difficult or nearly impossible to obtain precise information to aid selection of project delivery methods due to insufficient data and incomplete knowledge in the early stages of a construction project (Kim 2011; Weber et al.

2012). Fuzzy set theory is a commonly used tool in engineering to accommodate the involvement of subjective judgements, uncertainty, and imprecise data by utilizing membership functions and providing a possibilistic range of results (Bezdek 1993; Corona-Suárez et al. 2014; Kumar et al. 2013). Fuzzy Bayesian approach is able to accommodate and quantify both numeric and subjective variables in the decision-making process of selecting project delivery methods (Cheng et al. 2019; Eleye-Datubo et al. 2008; Ersel and İçen 2014; Ren et al. 2009; Sun et al. 2018; Yazdi and Kabir 2017). This approach has been adopted in many fields (e.g., engineering, business, economics, and computer science) to enhance the probability updating process with fuzzy evidences by utilizing the conditional probability densities and the membership functions of the evidence's values (Ferreira and Borenstein 2012; Kawamura 1993; Nguyen et al. 2016; Ung 2018; Zarei et al. 2019; Zhang et al. 2015; Zoullouti et al. 2017).

This chapter extends the work from Chapters 2 and 3 of this dissertation regarding the development of a fuzzy-based hybrid approach to support selection of project delivery methods in highway construction. Chapter 2 established seven highway project clusters that share high commonalities in project characteristics, project complexity, delivery risks, cost growth, and delivery methods using fuzzy cluster analysis. Chapter 3 developed a fuzzy rule-based inference system to recognize possible delivery methods of new highway projects based on historically established clusters using fuzzy pattern recognition. The fuzzy rule-based inferential engines are derived from fuzzy Gaussian membership function and reasoning based on the possibility theory (e.g., project A has 80% of chance to be recognized with D-B-B and 20% of chance to be recognized with D-B). This chapter proposes a hybrid approach of Bayesian theory and fuzzy set theory in terms of a fuzzy Bayesian inference system (FBIS) to support decision-making in selection of delivery methods for new highway projects.

RESEARCH MOTIVATION

Many studies have developed decision-aid models and frameworks to quantify risks and take into account uncertainty in the decision-making process (Molenaar and Songer 1998; Tran and Molenaar 2015). However, selection of project delivery methods also includes a large number of qualitative inputs that are difficult to be evaluated using current quantitative approaches (Gransberg and Shane 2010; Touran et al. 2011). This involvement makes the interaction between variables become even more complex in the selection process. This gap requires a scientific approach that can accommodate a combination of quantitative and qualitative variables to take into account uncertainty and risks in the selection process and the causality and interactions between selection criteria and project delivery methods.

Fuzzy set theory departures from traditional probabilistic approaches to leverage the evaluation of both quantitative and qualitative data in decision-making processes. Fuzzy sets, a mathematical approach developed by Zadeh (1978), have been used to mainly capture the qualitative inputs to generate deterministic decision-making models and widely applied to many engineering areas, such as computer science, mechanical engineering, aerospace engineering, and chemical engineering (Chan et al. 2014; Elwood 2014; Kruse et al. 2007). As a promising method to mathematically handle qualitative variables (Khanzadi et al. 2016), fuzzy sets were applied in this dissertation in terms of a soft clustering method, fuzzy cluster analysis, introduced by Bezdek (1993). The use of fuzzy sets requires extensive understandings of modern mathematics about human cognitive process and sophisticated designs (Li et al. 2013; Marzouk and Moselhi 2004). Therefore, fuzzy-based applications in construction often rely on particularly developed systems and models (Fayek and Lourenzutti 2018). However, the main limitation of fuzzy reasoning approaches is the lack of ability to conduct inference inversely (Li et al. 2012).

This limits the ability of investigating the causal relationships between project delivery methods and input project attributes.

RESEARCH OBJECTIVES

In the first two phases of this dissertation, a set of seven empirical project clusters and a fuzzy rule-based inference system were developed to help recognize the patterns of project delivery methods and cost growth used in highway construction. The developed fuzzy system experiences overlapping clusters, which leads to high testing error rates based on traditional probability theory. Therefore, the synergy of a commonly used approach in the probability theory, Bayesian networks, and fuzzy set theory is expected to overcome the shortfall of the developed fuzzy inference system and bridge the gaps in the current practices of project delivery method selection.

The objective of this chapter is to theoretically demonstrate a step-by-step procedure of using FBIS to support selection of project delivery methods in highway construction. In other words, how fuzzy-based variables are treated by Bayesian networks to increase the accuracy and reliability of the empirically developed rule-based inference system to identify a "true" delivery method for a new highway project. Given a set of project attributes of a new highway project, including project type, facility type, project size and duration, project complexity, and delivery risks, the theoretically developed FBIS is expected to identify an appropriate delivery method. This chapter also aims to scientifically accommodate a combination of qualitative and quantitative variables and take into account the interrelationship between them in the problem of selection of project delivery methods. Future work is to provide verification and validation of the applicability of the theoretical FBIS approach in supporting highway agencies in selecting appropriate project delivery methods.

LITERATURE REVIEW

This section describes a summary of current decision-aid approaches to the selection of project delivery methods and investigates the use of a hybrid approach of fuzzy sets and Bayesian networks in this area. Specifically, the practices of decision aids in selection of project delivery methods are summarized and discussed in terms of the evaluation of decision criteria. Subsequently, applications of fuzzy Bayesian approaches to construction decision-making is investigated to enhance the accuracy and reliability in project delivery method selection.

Evaluation of Decision Criteria in Selection of Project Delivery Methods

Selection of appropriate PDMs have been found to improve project performance, including lower cost growth, shorter schedule durations, higher quality, and better safety (Al Khalil 2002; Col Debella and Ries 2006; Ibbs et al. 2003). The traditional D-B-B delivery method is considered to foster adversarial relationships among project participants which often can result in negative performance outcomes (Park and Kwak 2017). On the other hand, alternative contracting methods (ACMs), including D-B and CM/GC, aim to shorten the project schedule, optimize total cost, and achieve a satisfactory level of project quality (Francom et al. 2016). Under particular circumstances, such as projects with a high level of uncertainty or complexity D-B has been found to provide better project performance than D-B-B (Nikou Goftar et al. 2014; Rojas and Kell 2008). In the U.S., state departments of transportation (DOTs) have increasingly used ACMs, which inspires the assessment of whether transportation projects result in better project performance and identify common performance patterns to support selection of PDMs (Touran et al. 2009b).

Decision-making frameworks for selection of appropriate PDMs are considered to be an important element of successful construction projects (Bakht and El-Diraby 2015; Mahdi and Alreshaid 2005). Many decision-making studies have been proposed to support project owners to select the most suitable PDM for their projects (Al Khalil 2002; Bypaneni 2017; Mostafavi and Karamouz 2010; Tran and Molenaar 2015). The typical methods of analysis in PDM decision-aid systems or models can be classified into two main categories: qualitative and quantitative approaches (Konchar and Sanvido 1998; Touran et al. 2011). The majority of current selection approaches in the construction industry are based upon subjective assessments of experts and guidelines from professional organizations (WSDOT 2016). Rising approaches of mathematical theories have been developed with critical selection criteria, mostly focusing on project performance. There are several types of inputs to decision-making models for PDMs including project characteristics, complexity, and project risks. Among all, project risks are considered as one of the most difficult inputs to quantify due to their subjective nature and qualitative units of measure (Diab et al. 2012). Project risks in selection of PDM are considered based on simulations and probabilistic approaches with which potential ranges of risk impacts to project performance are provided (Tran et al. 2016). However, it is difficult to incorporate two different types of inputs, numeric data from project characteristics and subjective data from risks, into a PDM-selection-supportive models. Thus, there is a need to utilize a new method to take into account both types of inputs and provide more accurate outcomes to support selection of PDMs.

Applications of Fuzzy Sets and Bayesian Networks in Construction Decision-Making

The main function of fuzzy set theory is to convert linguistic statements, which are only meaningful to human beings, to be quantifiable by a computer (Zadeh 1965, 1975). Fuzzy set theory has been used in a wide range of domains to evaluate incomplete, imprecise, or subjective inputs (Anderberg 2014). In the field of engineering, fuzzy set theory has been used to capture qualitative domain professional judgements to generate theoretical decision-making models and widely applied to many areas, such as computer science, mechanical engineering, aerospace engineering, chemical engineering, structural engineering, and construction management (Elwood 2014; D'Urso 2007; Hoppner et al. 1999; Nguyen et al. 2020; Seo et al. 2004). Within the construction industry, fuzzy set theory has been rarely used, and has mostly dealt with problems in risk-based management (Dikmen et al. 2007; Elbarkouky et al. 2016; Lam et al. 2001; Pawan and Lorterapong 2016). Hegazy and Ayed (1998) indicated that fuzzy set theory is notably useful in the construction industry where realistic historical project data are limited. In fact, it has demonstrated its applicability in quantifying some project performance metrics. For example, Knight and Fayek (2002) implemented fuzzy set theory and fuzzy logic to predict cost overruns of the design phase in vertical building projects. Li et al. (2006) attempted to forecast project status in terms of cost overruns and schedule delays based on fuzzy logic. Due to the dynamic nature of construction project data and information, the probability and severity of events cannot be satisfied by the crisp values (Ross 2010). Therefore, the probability of linguistic expressions can be transformed into fuzzy numbers.

Bayesian networks are an inference engine used to compute the probability of an event given the occurrences of other events. To construct a Bayesian network, a directed acyclic graph, consisting of nodes (i.e., discrete or continuous variables) and arcs (i.e., dependence between variables), is developed to illustrate the conditional dependence and causal relationships between nodes (Sun et al. 2018). The directed acyclic graph produces a set of conditional probabilities while missing arcs implies conditional independence between variables. The graph also allows joint and priori probability distributions to be specified and updated in terms of conditional probability tables (Zhang et al. 2016).

Fuzzy Bayesian inference is a hybrid technique that synergies a combination of possibility theory (fuzzy logic) and Bayesian networks (probability theory) in a unified inference system (Viertl 1987). It is defined as a multi-criteria decision-aid system which allows the decision maker to prioritize and select the most appropriate alternative under uncertainty (Sedki et al. 2010). By accommodating the causal interactions among the variables, FBIS allows the decision maker to solve more complex problems based on a variety of variables.

FBIS has been widely used in engineering research (Eleye-Datubo et al. 2008; Luque and Straub 2019; Zhang et al 2016; Uusitalo 2007), but limited studies have investigated its applications in project delivery decision problems. Based on the fuzzy rule-based inference system in Chapter 3, which was established using quantitative and qualitative variables, the proposed FBIS attempts to demonstrate the causal relationships between input variables (e.g., project characteristics, project complexity, delivery risks, and project cost performance) and project delivery method candidates. Using FBIS is able to provide understandable interpretations of variables compared to other machine learning approaches, such as support vector machines, decision trees, Monte Carlo simulation, sensitivity analysis, stochastic approaches, and artificial neural network (Guidotti et al. 2018; Ren et al. 2009). In addition, FBIS is capable of integrating different sources of information and knowledge, dealing with incomplete data sets, modeling causal relationships among variables, and accommodating uncertainty occurred in decision-making processes (Uusitalo 2007)

RESEARCH QUESTIONS

The motivation of this chapter is to overcome the shortage of the developed fuzzy inference system in Chapter 3 by incorporating a probabilistic aspect of Bayesian networks to increase the accuracy and reliability of the rule-based inference system with updatable conditional probability of variables. To achieve the research objective, this chapter attempts to address the following research questions:

1) *How do project attributes, project complexity, delivery risk factors, and cost performance interact with different project delivery methods (i.e., D-B-B and D-B)?*

    To answer the first question an application of Bayesian networks and fuzzy sets is proposed to describe the dependencies between variables both qualitatively and quantitatively. Causal relationships between facility type, project type, project complexity, delivery risks, cost performance, and project delivery methods are addressed in this research question.

2) *What new information would be gained by using a developed fuzzy Bayesian inference system to selection of project delivery methods?*

To answer the second question, a theoretical fuzzy Bayesian rule-based inference system is developed and expected to practically demonstrate via a Case example with a step-by-step procedure of selecting an appropriate delivery method associated with cost growth for highway projects.

THEORETICAL FRAMEWORK OF FUZZY BAYESIAN INFERENCE SYSTEM

In engineering decision-making, FBIS has been constructed by using two common approaches. The first approach develops a fuzzy Bayesian network with structures and parameters on the basis of historical data (Sun et al. 2018). This approach requires a large amount of training data which is often difficult to collect in the context of project delivery methods. The second approach establishes a fuzzy Bayesian network based on experts' experience and knowledge, which is commonly used in fuzzy Bayesian studies in engineering (Leu and Chang 2013; Nguyen et al. 2016). This chapter proposes the application of the first approach with empirical project data collected from highway agencies.

The key feature of the fuzzy Bayesian networks and inference system is to establish a possibility-probability directed acyclic graph of all variables to capture the logical network and provide causal probabilistic relationships between them (İçen and Ersel 2019; Islam and Nepal 2016; Liu et al. 2013; Viertl and Sunanta 2013; Zarikas et al. 2014). The inputs to the Bayesian Networks are fuzzy values determined by fuzzy membership functions. Figure 1 shows the entire process to develop a FBIS including the following steps:

- Determine empirical project clusters based on the commonality in project attributes, delivery methods, and cost growth between highway projects.
- Establish a fuzzy rule-based inference system to identify potential project delivery methods based on common groups of highway projects in terms of facility type, project type, project complexity, delivery risks, delivery methods, and cost growth.
- Create causal relationships between fuzzy values of project attributes, cost performance, and project delivery methods with probabilistic nodes and arrows from Bayesian networks. To extend the classic Bayesian networks into a fuzzy Bayesian inference system, it is essential to choose a proper fuzzy probability measure to conduct fuzzy Bayesian inference.
- Establish fuzzy Bayesian rule-based inference engines with prior, joint, and conditional probabilities observed and calculated from the previously developed fuzzy inference system. The final conditional probability table provides the degree to which a more appropriate project delivery method can be selected with updated information as the project is developing. The decision nodes (i.e., D-B-B and D-B) are identified in terms of True (%) and False (%) given a set of input variables and updated information as the project is developing.
- Conduct the defuzzification process to produce crisp values of the output project delivery methods. Validation and verification procedures are conducted to ensure the accuracy and reliability of the developed FBIS.

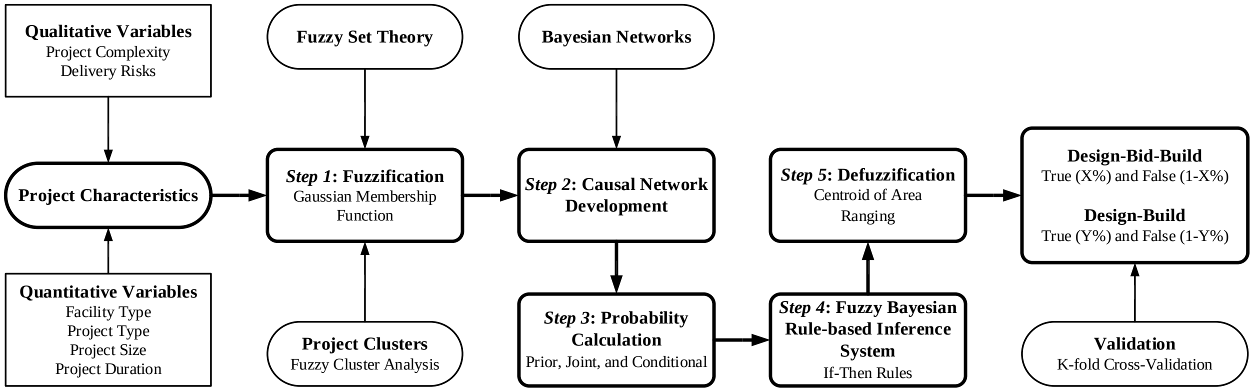

Figure 1. Research Approach of Fuzzy Bayesian Rule-based Inference System

The theoretical framework of FBIS in this chapter follows five specific steps: (1) fuzzify all input variables, (2) develop a causal (Bayesian) network including all variables or nodes (i.e., root nodes, intermediate nodes, and consequence nodes), (3) calculate all probabilities, including prior (or marginal), joint, and conditional, for the final Conditional Probability Tables of the developed network, (4) determine the fuzzy rule-based inference engines and select decision alternatives based on causal relationships between variable nodes, and (5) defuzzify the outputs of the fuzzy-Bayesian inference process.

The theoretical framework of FBIS in this chapter is developed based on a dataset X consisting of n highway projects, $X = \{X_1, X_2, X_3, ..., X_n\}$. Each project X_i has m variables (i.e., project characteristics, project complexity, delivery risks, and project performance indicators) $X_i = \{X_{i1}, X_{i2}, X_{i3}, ..., X_{im}\}$. Using fuzzy cluster analysis, a set of c predefined clusters serve as classifiers in the fuzzy rule-based inference system $C_j = \{1, 2, 3, ..., k\}$. Prior to inputting data to the fuzzy Bayesian networks, a standardization process is conducted to ensure that all of the variables have a unified scale to avoid impacts of dissimilarity measures. Ranging method is one of the recommended standardization methods in the fuzzy logic area and denoted as Equation 1 (Klir and Yuan 1995):

$$X_{STD} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

Where: X_{STD} = standardized value; X = crisp value; X_{min} = minimum value of the project attribute; X_{max} = maximum value of the project attribute.

<u>Step 1</u>: Fuzzification of Inputs

The fuzzifying process converts input values into standardized fuzzy numbers prior to the inferential reasoning process. The result of this process provides a set of standardized input variables (quantitative and qualitative) in a unified scale to avoid impacts of dissimilar measures. Each input project is represented as a multi-dimensional data vector, consisting of the number of selected variables. The detailed fuzzification process can be found in Chapter 3: Fuzzy Pattern Recognition. The objective of the fuzzification process is to manipulate input data with Gaussian fuzzy membership function to convert raw inputs and insert membership values (Hwang 2004). The Gaussian functions are more appropriate than other types of membership functions, including triangular, trapezoidal, and sigmoid to handle a combination of quantitative and qualitative inputs (Daneshvar 2011). The Gaussian membership function has the maximum value of 1. The linguistic labels are assigned to each approximated membership function of input variables based on the values of cluster centers. The Gaussian membership function is illustrated in Equation 2 (Ross 2010):

$$f_i(x;\mu,\sigma) = e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{2}$$

Where μ_i and σ_i are the cluster center and spread of the i^{th} project attribute with Gaussian membership function f_i (i.e., facility, project type, project size and duration, project complexity, delivery risks, and cost growth). Detailed of the establishment of cluster centers and spreads can be found in Nguyen et al. (2020).

Step 2: Development of Bayesian Networks (Causal Relationships)

Bayesian networks contribute to the establishment of causal relationships between pre-defined clusters, project size and duration, and delivery method candidates (i.e., D-B-B and D-B). Figure 2 shows the Bayesian network of the 17 selected variables in the fuzzy rule-based inference system established in Chapter 3. The seven project clusters were developed based on project characteristics, level of project complexity, delivery risk factors, and historical cost growth in Chapter 2. The seven clusters and input variables of project size and duration serve as parent nodes to help estimate the posterior conditional probability of project delivery methods (i.e., D-B-B and D-B). The final results of two delivery methods, considered as discrete variables, are represented in two states: *True* and *False*, associated with a statistically probability-possibility percentage. The established Bayesian network between input and output variables is a foundation to determining fuzzy Bayesian inference engines in the next step.

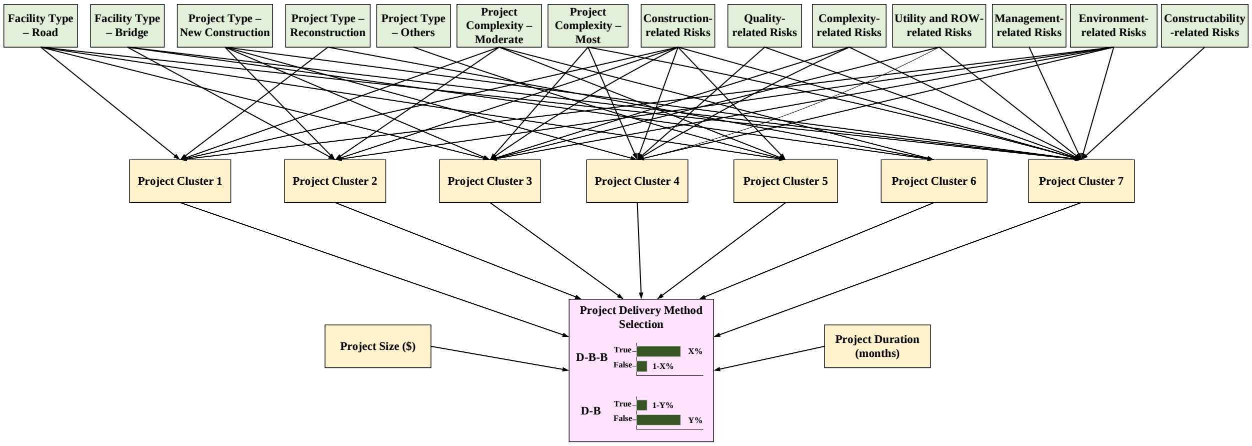

Figure 2. Development of Fuzzy Bayesian Networks

Figures 3 and 4 illustrate the causal relationship between (1) project attributes and project clusters and (2) project clusters, project size, project duration, and delivery methods (i.e., D-B-B and D-B), respectively. In Figure 3, the relationships between project attributes (i.e., project type, facility type, project complexity, and delivery risk factors) and project clusters were developed based on fuzzy clustering analysis. In Figure 4, in addition to seven project clusters, project size and duration are added to increase the accuracy of the selection of project delivery methods. Figure 5 illustrates an example of identifying project delivery methods using project cluster 1, where the conditional probabilities of D-B-B and D-B in terms of “True” or “False” percentages are provided. Specifically, the conditional probabilities of project cluster 1 are updated based on the prior probabilities of input project attributes, including facility type – road, project type – reconstruction, moderate project complexity, construction risk factor, and environment risk factor.

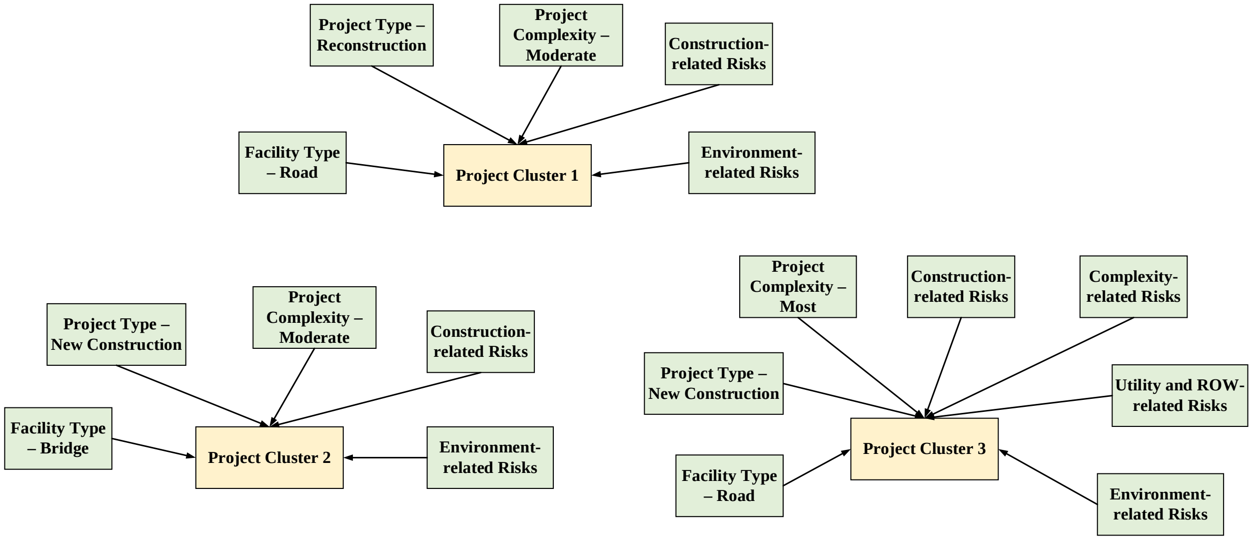

Figure 3. Causal Relationships between Project Attributes and Empirical Project Clusters

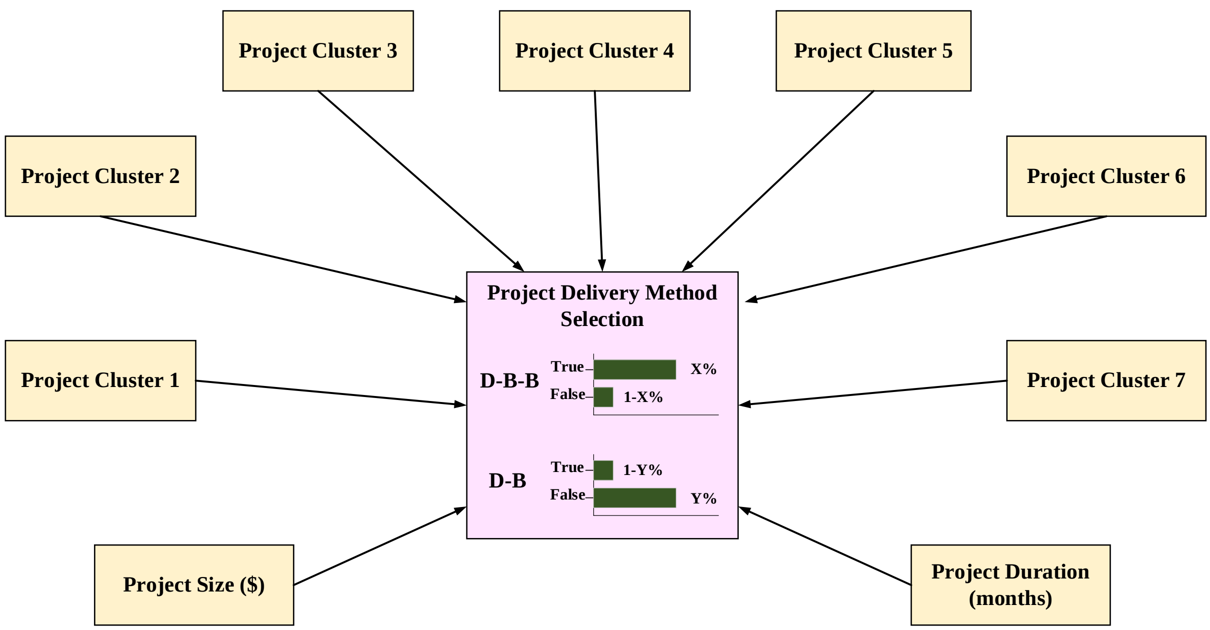

Figure 4. Causal Relationships between Project Clusters, Project Size and Duration, and Delivery Methods.

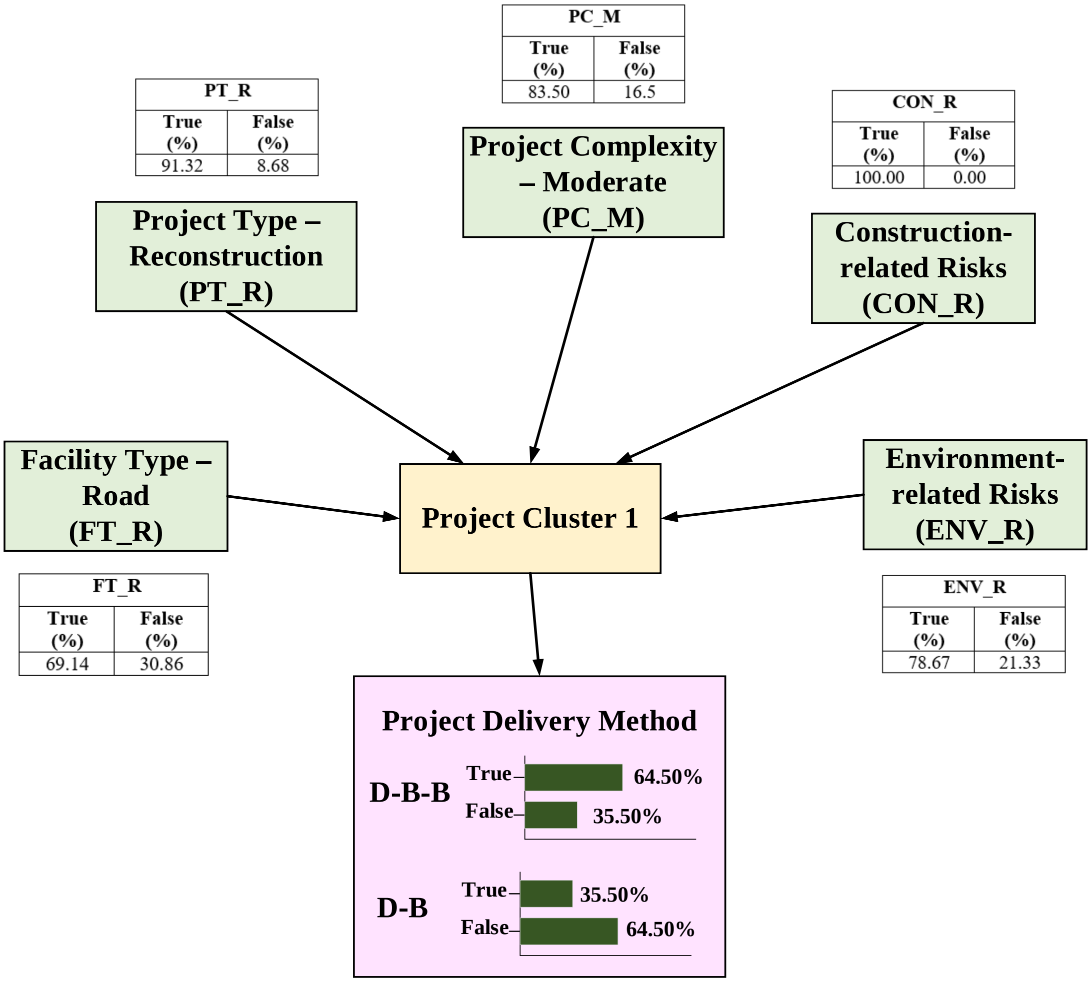

Figure 5. Example of Identifying Project Delivery Methods Using Project Cluster 1

<u>Step 3</u>: Calculation of Probabilities for Conditional Probability Tables

The developed directed acyclic graph shown in Figure 2 is used to represent the probabilistic causal relationship between the historical project characteristics, clustered project groups, new project characteristic inputs (cost and duration), and project delivery method candidates (i.e., D-B-B and D-B). The graphical conditional dependencies between facility type, project type, project size and duration, project complexity, delivery risks, and cost growth are probabilistically represented using the conditional probability tables. Three typical probabilities, including priori, joint, and conditional, of the developed FBIS are calculated and updated based on Bayes theorem (Corona-Suárez et al. 2014; Sedki et al. 2010).

First, the joint probability of the input project attribute X is denoted as Equation 3:

$$P(X) = \sum_{i}^{n-1} P(X_i \mid X_{i+1}, \ldots, X_n) \tag{3}$$

Where: X_{i+1} is a causally related project attribute of X_i. The probability of project attribute X in the i^{th} project $P(X_i)$ is calculated as $P(X_i) = \sum_{\frac{X}{X_i}} P(X)$.

Second, the prior probability of the input project attributes is calculated as follows:

- Facility type includes road, bridge, drainage, intelligent transportation system (ITS), and other. The prior probability of each sub-category is computed based on the frequency of them in the collected dataset. For example, the probability of road variable equals to the percentage of road projects throughout the entire dataset (i.e., if road represents 60% of the collected highway projects, the prior probability of road variable is 0.6).

- Project type includes new construction, reconstruction, and other. Similar to the facility type categories, each sub-project type produces prior probability based on the percentage of them throughout the entire dataset.
- Project complexity produces the prior probability of each level of complexity including the most, moderate, and none complex with similar computation to the previous variables.
- Delivery risk factors 1 to 7 produce the prior probability by calculating the frequency of each delivery risk and multiplied by the exploratory factor loadings within each risk factor. For example, Risk Factor 3 – Constructability consists of two delivery risks, including Risk #17: delays in procuring critical materials, labor, and specialized equipment with a factor loading of 0.85 and Risk #18: significant increase in material, labor and equipment cost with a factor loading of 0.76. The prior probability of Risk Factor 3 is calculated based on the prior probability of risks #17 and #18 throughout the entire dataset.
- Cost growth produces prior probability of five categories: savings, none, low, medium, and high with similar computation to the previous variables.

As the project is thoroughly developing, newly updated information of input variables can be loaded to the FBIS to enhance the accuracy of the selection of an appropriate delivery method. The priori probabilities can be updated by using Equation 4 below (Kim 2011; Islam and Nepal 2016; Yang et al. 2006):

$$P(X_i \mid T_j) = \frac{P(X_i \cap T_j)}{P(T_j)} = \frac{P(X_i \cap T_j)}{\sum_{X_i} P(X_i \cap T_j)} = \frac{P(T_j \mid X_i) \times P(X_i)}{P(T_j)} \tag{4}$$

Where:

T_j represents a set of new information given throughout the project development that affect the created variables in the FBIS.

$P(X_i | T_j)$ is the posterior probability of variable X_i given new information T_j

$P(X_i)$ is the prior probability of variable X_i calculated from the data-driven clustering groups.

$P(T_j)$ is the probability of state j of the newly given information T, which can be estimated from the developed Bayesian networks.

$P(T_j | X_i)$ is the degree of belief in the accuracy of variable X_i given the new information T_j, which can be calculated based on the fuzzy rule-based inference system in Chapter 3.

Third, the conditional probability table of all the nodes (e.g., root, intermediate, and consequence) in the proposed Bayesian network is computed with fuzzy values derived from Gaussian membership function as shown in Table 1. The root nodes include project attributes: project type, facility type, project complexity, delivery risks, and historical cost growth. The intermediate nodes consist of empirically identified project clusters, project size, and project duration. The consequence (or leaf) nodes consist of two project delivery method candidates: D-B-B and D-B.

Table 1. Theoretical Conditional Probability Table

Variable	Project Delivery Method Candidates (PDM)	
$P(PDM \mid C_i, m_i, x_i)$	Design-Bid-Build	Design-Build
$P(C_i \mid m_i, x_i)$	Seven Project Clusters, Project Size, and Project Duration	
$P(m_i \mid x_i)$	Facility Types, Project Types, Project Complexity, Delivery Risk Factors	
$P(x_i)$	Road, Bridge, New Construction, Re-construction, Moderate Complex, Most Complex, Risk Factor 1, Risk Factor 2, Risk Factor 3, Risk Factor 4, Risk Factor 5, Risk Factor 6, Risk Factor 7	

<u>Step 4</u>: Determination of Inference Engines and Decision Alternatives

Mapping the fuzzy input data into the Bayesian belief network is the process of determining the degree to which the project cluster C_i belong to a project delivery method (i.e., D-B-B or D-B) associated with conditional probabilities of the causal relationship between them. A rule base is projected from the membership function and written in terms of the input characteristics and group labels (rule antecedents and rule consequents, respectively) (Elwood 2014). The established rule base is able to accommodate the fuzziness and membership values of variables. This chapter continues to utilize the definition of Mamdani-type fuzzy inference system developed in Chapter 3 to logically reason multiple inputs via "If-Then" rule-based engines and produce multiple outputs. Specifically, an example of rule-based inference engines is shown below:

IF Cluster 1 is {%True and %False} and Cluster 2 is {%True and %False } and Cluster 3 is {%True and %False } and … Cluster 7 is {%True and %False } and Project Size is {small, medium, or high} and Project Duration is {decreased or increased} THEN the appropriate project delivery method is D-B-B {%True and %False} or D-B {%True and %False}.

Where: Clusters 1 to 7, Project Size, and Project Duration are mutually different fuzzy sets while *{%True and %False}* is calculated with the conditional probability tables.

If-then propositions of this form can be aggregated (the process of determining the final consequent over all of the rules) in different ways (D'Urso 2007). In the Max-Min inference method, the "and" conjunction in each rule is evaluated by the fuzzy set minimum operator and the aggregation over all the rules is evaluated by the maximum operator (Elwood 2014).

The probability results from the conditional probability tables provide a rule-based probability-possibility inferenced outcome of each project delivery method to investigate how suitable of them to particular project input attributes. Since the outputs are represented in terms of fuzzy values, the output with the maximum Gaussian membership function is selected. Specifically, the outcomes of the proposed FBIS show the following:

- Suitability of D-B-B = True Percentage = $P(D\text{-}B\text{-}B \mid C_i, m_i, x_i) = X\%$
- Unsuitability of D-B-B = False Percentage = $P(D\text{-}B\text{-}B \mid C_i, m_i, x_i) = 1\text{-}\ X\%$
- Suitability of D-B = True Percentage = $P(D\text{-}B \mid C_i, m_i, x_i) = Y\%$
- Unsuitability of D-B = False Percentage = $P(D\text{-}B \mid C_i, m_i, x_i) = 1\text{-}\ Y\%$

Step 5: Defuzzification of Outputs

The inferenced outcomes of decision alternatives are hardened to convert fuzzy outcomes into crisp values for subsequent humanistic judgements; this process is called defuzzification. Some common defuzzification approaches in engineering applications include the centroid of area, maxima, mean of maxima, and center average weighting (Ung 2018). This chapter utilizes the centroid of area approach to determine the outcome of the proposed FBIS. Equation 5 shows the calculation of the centroid of area approach (Zhang et al. 2016):

$$Centroid_{PDM} = \frac{\int x_i \times \mu_{PDM}(x_i) x_i dx_i}{\int x_i \times \mu_{PDM}(x_i) dx_i} \tag{5}$$

Where μ_{PDM} represents the maximum membership value of the Gaussian membership function *x* of the i^{th} input project attribute.

Although this approach requires a high load of computing burdens when using Gaussian membership function, it is capable of providing highly accurate outcomes (Ross 2010). Figure 6 shows the example of using the centroid of area approach to determine the maximum Gaussian membership value of D-B-B with the state of being *True* = *X%*.

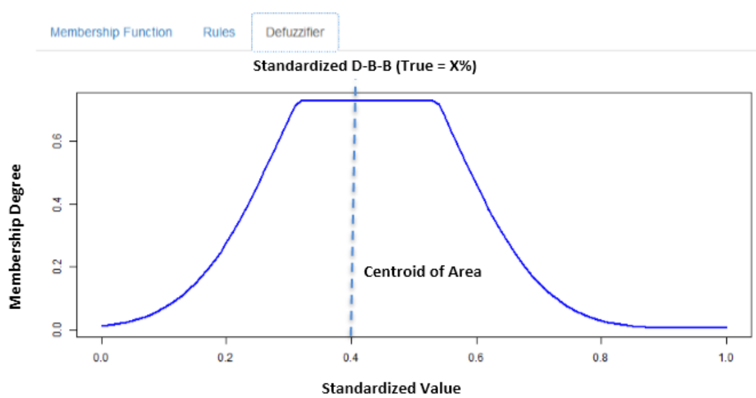

Figure 6. Defuzzification of FBIS using Centroid of Area Approach

In addition, ranging method is utilized to maximize the interpretability of the fuzzy rule-based inference system (Kruse 2007). Equation 6 is used to transform the fuzzy values to crisp values of the FBIS's outcomes (Nguyen et al. 2020).

$$X = X_{STD} \times (X_{\max} - X_{\min}) + X_{\min} \tag{6}$$

Where X_{STD} = standardized value; X = crisp value; X_{min} = minimum value of the project attribute; X_{max} = maximum value of the project attribute.

VALIDATION

To validate the proposed fuzzy Bayesian inference system, a 10-fold cross-validation is proposed to determine the validation and testing error rates. This cross-validation approach is able to produce low bias and low variance results (Sargent 2013). Using this validation approach, the data is divided into two main sets: training and testing; while the training set also includes the validation set used in the subsequent cross-validating process. For fuzzy-based classification problems, a 90:10 split ratio is recommended to treat the dataset (Ross 2010). In other words, 90% of the entire dataset is used for training the proposed FBIS while 10% of the entire dataset is used for testing the system. Specifically, the training set is split into ten subsets (i.e., ten folds) of equal size. The nine subsets are used to train the system while the remaining one subset is used to validate the system with ten iteration rounds of re-sampling the entire training set. The testing set is used separately to assess the accuracy of the FBIS's outcomes. To apply FBIS in practice, a web-based user interface can be developed with the R programming language, which allows users to input numerical project data, including facility type and project type, along with qualitative judgements, including project complexity and delivery risk factors, to provide predictive cost growths associated with appropriate project delivery methods.

DISCUSSION

This section summarizes the theoretical implications of the theoretically developed fuzzy Bayesian rule-based inference system within the domain of project delivery method selection. These implications are expected to shed light on the use of Bayesian networks to overcome shortages of fuzzy set theory to support project delivery decision-making under uncertainty.

Theoretical Implications

The theoretical FBIS framework provides decision makers with insights regarding the implementation of Bayesian networks to (1) establishing the structural relationships between project delivery decision variables and (2) increasing the accuracy of fuzzy input variables using prior and posterior probabilities. These two insights are expected to help decision makers overcome the constraints of fuzzy-based applications.

Due to the fact that fuzzy cluster analysis increases the overlaps among clustered patterns, the outcome of fuzzy-based classification often reserves pretty high training and testing error rates compared with normal statistical tests (Elwood and Corotis 2015, Ross 2010). By computing and updating the prior and posterior probabilities of fuzzy input variables, Bayesian networks are capable of improving the interactions between them and supporting fuzzy rule-based inference engines. This chapter recommends the use of R programming to create the FBIS framework with two packages of *BnLearn* and *gRain*. Details of these two R packages can be found in Scutari (2009).

Structural Relationship between Variables in Project Delivery Method Selection

The first implication refers to the theoretically developed Bayesian (causal) network between project attributes, seven empirically established project clusters, and project delivery methods. This network can be used for determining the existing statistical dependencies between variables in project delivery method selection. The structure of the Bayesian network in this dissertation was formed based on statistical evidences (the seven pre-defined project clusters); thus, this is defined as a Bayesian probability network (Zarikas 2014).

The relationship between the seven pre-defined project clusters and input project attributes can be identified based on the within- and between-clusters membership values of those attributes. For instance, in the project cluster 1, reconstruction road projects were correlated with the moderate project complexity and two delivery risk factors (construction and environment risks). In the project cluster 2, new construction bridges were also correlated with the moderate project complexity and two delivery risk factors (construction and environment risks). In project cluster 3, new construction roads were correlated with the most project complexity and four delivery risk factors (construction, complexity, utility-and-ROW, and environment risks). In the project cluster 4, new road construction projects were correlated with the most project complexity and five delivery risk factors (construction, quality, complexity, utility-and-ROW, and environment risks). In the project cluster 5, other types of road projects, such as road widening and new sidewalks, were correlated with the moderate project complexity and one delivery risk factor (construction risks). In the project cluster 6, reconstruction bridges were not correlated with any critical delivery risk factors, but the moderate project complexity. In the project cluster 7, new construction roads and bridges were correlated with the most project complexity and all seven delivery risk factors.

Highway agencies can utilize the outcome of the theoretical Bayesian probability network to identify the interrelationship between project characteristics, impacts of delivery risks, level of project complexity, and cost performance and select an appropriate delivery method for their new highway projects. Because the seven project clusters were established based on the degree of the commonalities across input project attributes, there is a need of collecting a large data sample size to improve the validity and reliability of the developed Bayesian probability network.

Computing and Updating Probabilities of Input Variables

Another theoretical implication to project delivery decision makers is that using Bayesian probability networks can improve the accuracy of fuzzy input variables by computing and updating prior and posterior probabilities. The prior probabilities of the input fuzzy variables were calculated based on the likelihood of their occurrence (i.e., fuzzy membership values) within a particular cluster, which might be affected by the overlaps between fuzzy-based project clusters. However, the posterior probabilities were computed based on Bayes theorem taking into account the probabilistic causal relationship between the variables and updated project information (İçen and Ersel 2019). This process provides a legitimate set of conditional probability distributions used in the rule-based inference engines to reason the suitability of project delivery methods. The final conditional probability table produces the specific *True* and *False* probabilities of each project delivery method candidate. This outcome is expected to help public agencies determine the most appropriate delivery method for their new highway projects given the initial project attributes and updated project information.

CONCLUSION

The majority of decision-making in selection of project delivery methods in highway construction rely upon a combination of qualitative (e.g., experts' experience and judgements on delivery risk impacts and levels of complexity) and quantitative (e.g., project size, duration, type, and historical performance) variables. This chapter attempts to quantify a set of qualitative and quantitative variables to support decision-making in selecting an appropriate project delivery method for new highway construction projects. A deterministic probability-possibility inference system with a synergy of fuzzy set theory and Bayesian causal networks was proposed to aid the selection of project delivery methods in highway construction. This inference system is anticipated to overcome the drawbacks in the established fuzzy rule-based inference system in Chapter 3. An empirical dataset of 254 highway projects is used to validate and verify the developed theoretical framework. A Case project is selected to illustrate and test the system. Validation or applicability of fuzzy Bayesian rule-based inference system in highway project delivery decision-making by cross-validation approach and a Case project.

This chapter is anticipated to contribute to the body of knowledge by proposing a theoretical framework of FBIS with a step-by-step process to identify an appropriate delivery method for a new highway project based on associated project characteristics, level of complexity, delivery risks, and historical clusters. The developed system provides a comprehensive understanding of causal relationships between facility type, project type, project complexity, delivery risks, cost performance, and project delivery methods. Highway agencies can use the developed system as a reference to select an appropriate delivery method for new projects based on historic data.

Limitation and Future Work

There are several limitations in the theoretically developed FBIS framework that can be addressed in future work. First, the number of project clusters established was limited due to the availability of historical highway project data. In fact, the use of fuzzy sets enables a high degree of overlaps between project clusters based on the membership coefficients of each project within a given cluster (Kassambara 2017). The distance between project data and the cluster center determines the possible number of project clusters. Therefore, a larger data sample size can reduce the total overlapping variation and increase the compactness of the project clusters.

Second, delivery risks are one of the critical variables in any decision-making processes of project delivery method selection (Tran and Molenaar 2015). This qualitative variable involves a high degree of subjective judgements and depends upon experience and risk preferences (i.e., risk averse, risk neutral, and risk acceptant) of decision makers. The risk ratings of the pre-defined project clusters were provided when the projects were done, which might expose to some biases. Future work should consider collecting delivery risk assessments that were completed prior to project initiation. In addition, the seven pre-defined project clusters accounted for only delivery risk impacts on cost performance. The potential delays due to schedule risk events are also critical in the selection of project delivery methods (FHWA 2018). Thus, schedule risk ratings should be also considered in future work to provide decision makers with insights regarding the interrelationship between cost and schedule delivery risks.

Third, other project characteristics, such as procurement and payment methods, are interconnected with project delivery methods. The procurement process and payment provision can also be empirically clustered to help recognize the common combinations of those methods associated with particularly selected project delivery methods. In addition, the total value of the change/extra work orders also greatly contributes to the selection of project delivery methods. This decision variable may include agency directed changes, changes in planned quantities, unforeseen or external project conditions, and errors and omissions in the plans.

CHAPTER 5:

CONCLUSION

SUMMARY

Decision-making in construction projects is highly complex and involves numerous risks and points of uncertainty because of the dynamic nature of the construction industry. To make an appropriate decision, a set of relevant decision variables are necessarily identified and evaluated under uncertainty. Within the context of a construction project, the relevant decision variables often include both quantitative and qualitative factors. This dissertation attempted to support decision-making processes under uncertainty in construction projects by demonstrating the applications of fuzzy set theory and fuzzy logic, which are scientific and quantitative approaches to model and measure a combination of qualitative and quantitative decision variables. Accordingly, the utilization of fuzzy-based approaches was illustrated using the decision problem of project delivery method selections in highway construction projects.

Selection of project delivery methods is a rising problem of the current construction industry to pursue better project performance which requires more empirically scientific approaches. The majority of project delivery method selection in highway construction rely upon subjective judgements and experience of decision makers. The selection of project delivery methods typically involves two types of variables: quantitative (e.g., project type, facility type, project size, and project duration) and qualitative (e.g., project complexity and delivery risks). Current decision aids in project delivery method selection lack deterministic approaches to accommodate a combination of quantitative and qualitative variables. To bridge this gap, this dissertation aimed at supporting selection of project delivery methods by utilizing applications of fuzzy set theory and fuzzy logic to simultaneously take into account common quantitative and qualitative variables and empirically investigate the relationship between them. Three fuzzy-

based approaches were derived from fuzzy cluster analysis, fuzzy pattern recognition, and fuzzy Bayesian networks. The motivating question for this dissertation is shown below:

"How to demonstrate the applications of fuzzy set theory and fuzzy logic to support project delivery method selections in highway construction?"

To answer to the motivating question, this dissertation sampled an empirical dataset of 254 completed highway construction projects collected from 28 public agencies across the U.S. Then, common groups of projects within the collected data were identified using fuzzy cluster analysis. Next, the pre-defined groups were used to train and test the fuzzy inference system. Finally, Bayesian networks were incorporated to leverage the accuracy of fuzzy variables using prior and posterior probabilities.

Main Findings and Discussion

Within the context of project delivery method selections, this dissertation supports public agencies to empirically identify the common trends (patterns) of using delivery methods in their highway projects in terms of project complexity, delivery risks associated, and cost performance. Taking into account the empirically identified patterns of project delivery methods, this study developed a quantitative inference system to help the agencies match their new coming highway projects with their historical database. Accordingly, the use of delivery methods associated with cost growth in the previous projects that share high similarities in project characteristics, the level of complexity, and delivery risks involved can be used as a reference to make better project delivery decisions.

Demonstrating three fuzzy-based applications provides several findings. First, fuzzy cluster analysis was used to establish seven project clusters which produce critical observations related to the appropriateness of project delivery methods to project characteristics and risk profiles. Accordingly, new highway construction projects delivered by D-B was privileged over D-B-B in terms of more project complexity and higher delivery risk impacts which is in consistent with previous studies (CII 2018; Goodrum et al. 2011; Hale et al. 2009; Koppinen and Lahdenperä 2004; McWhirt 2007; Migliaccio 2006; Rosner et al. 2009; Shrestha 2007a, 2007b; West Valley Construction, 2019). D-B projects are found with mostly moderate complexity. However, this result is not in line with Minchin et al. (2013), which compared sixty highway and bridge projects from Florida DOT and stated that using D-B-B resulted in better cost performance than using D-B. A possible reason is that this study takes into account uncertainty impacts of project complexity and risks while Minchin et al. (2013) did not include these inherent factors.

Based on the sample of 254 projects, this study showed that highway projects delivered by D-B-B was preferred over D-B in terms of reconstruction projects with higher construction and environmental risks. New project, either road or bridge, are likely to have higher delivery risk impacts and the most complexity if procured by D-B; however, it only shows very low-cost growth with the mean of 0%. Reconstruction bridge projects with moderate complexity are likely to have no delivery risk impact if procured by either D-B-B or D-B. The construction and environmental risk factors are likely to have higher impacts on D-B projects than D-B-B projects. New construction projects have higher cost growth than reconstruction and other project types. Road projects have higher cost growth than bridge projects. To help practitioners investigate the differences between D-B-B and D-B in terms of facility type, project type, project complexity, associated delivery risk factors, and cost growth, pairwise comparisons between seven project clusters are provided as follows:

- *D-B-B versus D-B in Road Projects*

  Road projects delivered by D-B performed better than D-B-B road projects in terms of cost growth. This result is in line with Shrestha et al. (2012) and Tran et al. (2018), where a total of 4,203 highway projects delivered by D-B-B and D-B was analyzed. According to clusters 1 and 4, D-B was selected for new construction road projects with high complexity and risk impacts while D-B-B was selected for reconstruction road projects with lower complexity. Both clusters were observed with low cost growth and slightly different between D-B-B and D-B projects, 2% and 5%, respectively. In new construction roads (clusters 3, 4, and 7), projects procured by D-B were observed with lower cost growth than projects delivered by D-B-B even though D-B projects had higher complexity and risk impacts. Quality, constructability, and construction risks were observed to have a critical impact on cost performance in new D-B road projects. In reconstruction road projects (clusters 1 and 5), D-B-B was mainly selected to use with moderate project complexity and low risk impacts, which resulted in low cost growth. Environmental risks were observed to have a critical impact on cost performance when using D-B-B in reconstruction road projects.

- *D-B-B versus D-B in Bridge Projects*

  D-B-B reconstruction-bridge projects were observed to have lower cost growth than D-B delivered new-construction-bridge projects according to clusters 2, 6, and 7. This is not in line with other PDM-comparisons-based studies, such as Touran et al. (2009a, 2011) and Sullivan et al. (2017). A possible reason is that the D-B-B bridge projects in the dataset were reconstructed with only moderate complexity and observed with very low risk impacts resulting in 0% of cost growth. On the other hand, D-B was selected for new-construction-bridge projects with high complexity and risk impacts resulting in low cost growth. Construction risks were observed to have a critical impact on the use of D-B in this case.

- *D-B-B versus D-B in terms of Project Complexity and Risks*

  New construction highway projects with high complexity and risk impacts selected D-B instead of D-B-B and resulted in none-to-low cost growth. Under circumstances where D-B-B was selected for complex highway projects, medium-to-high cost growths were observed even though there were very low risk impacts involved. Contrarily, reconstruction highway projects with moderate complexity delivered by D-B-B were observed to have lower cost growth than D-B projects. Construction, utility and ROW, and environmental risks were observed to have a high impact on cost performance of D-B-B highway projects while D-B projects experienced high impacts of construction risks.

Second, based on the seven established project clusters, fuzzy pattern recognition was used to develop the fuzzy rule-based inference system which systematically maps the inputs of a new highway project to one or more project clusters (i.e., fuzzy classifiers) based on the similarity in project attributes (i.e., membership values within seven identified project clusters) and provides information regarding the use of delivery methods associated with cost growth. This inference system does not attempt to quantify any project delivery selection criteria, yet relationships between chosen project variables were established in terms of fuzzy membership functions to explore the similarities between them. Because of the inherent overlaps of project variables between fuzzy classifiers, a new project can belong to more than one cluster. Therefore, the representation of each fuzzy classifier relies on the control variable – type of delivery methods used and associated cost growth. To account for the qualitative input variables, different judgements on seven risk factors may lead to multiple decision scenarios with different patterns of project delivery methods associated with cost growth to aid the decision maker to select an appropriate delivery method. Therefore, in many cases, highway construction projects with similar facility types and project types may end up with different cost growths and delivery methods used.

A programming-based graphical user interface to support selection of delivery methods with new project inputs of characteristics, project complexity, and delivery risk impacts, was coded via R programming. This chapter also provides public agencies with seven common patterns of highway projects in terms of project delivery methods and associated cost growth. Utilizing fuzzy pattern recognition increases the overlaps among the clustering groups. Therefore, this might not be optimistic in terms of statistics and probability theory because of a great error rate. However, in the context of fuzzy set theory, it is acceptable because fuzzy sets are non-interactive, and there is restriction of a membership value based on assignment of other membership values (Elwood and Corotis 2015; Ross 2010).

Third, the theoretical framework of fuzzy Bayesian inference system aimed to overcome the restriction of fuzzy set theory by investigating the dependencies between project characteristics, project complexity, delivery risks, cost growth, and project delivery methods, as well as calculating and updating probabilities of all variables as new information becomes available. There are five steps to establish the theoretical framework, including:

1. Fuzzify all variables, including project type, facility type, project size and duration, project complexity, delivery risks, and cost growth.
2. Develop a Bayesian network including all variables.
3. Calculate all probabilities, such as prior, joint, and conditional, for the final Conditional Probability Tables of the developed network.
4. Determine the fuzzy rule-based inference engines and select decision alternatives based on causal relationships between variables.
5. Defuzzify the outputs of the fuzzy-Bayesian inference process.

RESEARCH CONTRIBUTIONS

This dissertation contributes to the body of knowledge by demonstrating the applications of fuzzy hybrid approaches to support decision-making under uncertainty within the decision scenario of project delivery method selections. A set of quantitative and qualitative decision variables in the selection of delivery methods for highway projects were modeled and assessed using fuzzy set theory and fuzzy logic. A theoretical framework of an empirical inference system utilizing fuzzy set theory and Bayesian networks was also proposed to enhance the accuracy of fuzzy-based applications in decision-making under uncertainty. This study is one of the first attempts that applied fuzzy cluster analysis, fuzzy pattern recognition, and fuzzy Bayesian inference system to support the selection of project delivery methods in highway construction. This dissertation also sheds light on the applicability of fuzzy set theory, fuzzy logic, and fuzzy hybrid approaches to the broader construction research areas within the topics of decision-making under uncertainty and the opportunity to utilize statistical techniques that are well-suited for such decisions.

To construction practitioners, the implications of this dissertation can benefit public highway agencies in making better project delivery decisions based on their historical data and inputs of project characteristics. In other words, given the information of a new highway project, the agency can utilize the developed fuzzy-based applications to identify an appropriate delivery method associated with a potential range of cost growth. In addition, they can also update their input information during the project development to improve the accuracy of their decision-making. This study provides several practical implications in project delivery method selections. The seven empirical groups of highway projects that share high commonalities in project attributes provided insights regarding cost performance comparisons between D-B-B and D-B delivery methods to highway agencies. Using the fuzzy rule-based inference system developed based upon the seven project clusters, new highway projects can be systematically matched with an empirically recognized delivery methods associated with potential cost growth. Decision makers can also adjust the inputs of delivery risk impacts and levels of complexity to match their risk attitudes and opinions about project complexity. A graphical user interface was developed by R programming to help decision makers recognize project delivery method patterns with the inputs of project characteristics, project complexity, and delivery risks. The theoretical framework of fuzzy Bayesian inference system provided a comprehensive understanding of causal relationships between facility type, project type, project complexity, delivery risks, cost performance, and project delivery methods.

LIMITATIONS

Although this dissertation is one of the first studies attempting to investigate the hybrid fuzzy approaches to accommodate a combination of quantitative and qualitative variables in the selection of project delivery methods, it has several limitations as follows:

1) **Increasing the sample size**. There was a lack of collected CM/GC-delivered highway projects (n=34) compared with D-B-B (n=114) and D-B (n=119) in the collected dataset. The collected cost risk ratings might be biased because of project administrators and managers were asked to rate the risk profiles when the project had already completed. Only five project attributes: facility type, project type, project complexity, delivery risks, and project delivery methods were considered. Cost growth was the only project performance indicator that was taken into account.
2) **Expanding the hybrid fuzzy approaches to other construction sectors**. This dissertation concentrated on highway construction projects. Therefore, the results would not be applicable for other construction sectors, such as vertical, medical, and aviation.
3) **Investigating the underlying effects across project characteristics, delivery risks, project delivery methods, and cost performance.** The fuzzy-based applications in this dissertation did not attempt to examine the underlying effects of facility type, project type, project size, project duration, project complexity, and delivery risks on cost growth associated with project delivery methods.
4) **Validating and testing the hybrid fuzzy approaches.** The validation of the developed fuzzy inference system in Chapter 3 experienced relatively high training and testing error rates compared with the normal statistical tests.

5) **Expanding the hybrid fuzzy approaches to other decision-making scenarios.** This dissertation only focused on supporting project delivery method selection.

POTENTIAL AREAS FOR FUTURE RESEARCH

The existing limitations in the application of fuzzy sets and hybrid fuzzy approaches can be removed with the suggestions for future research under three major areas.

- First, the limitations #1 and #2 can be removed by collecting more data, considering other decision variables, and expanding the developed hybrid fuzzy approaches to other sectors.
- Second, the limitations #3 and #4 can be removed by incorporating the developed hybrid fuzzy approaches with other statistical and machine learning techniques.
- Third, the limitation #5 can be removed by considering the application of the developed hybrid fuzzy approaches in other construction decision scenarios.

Fuzzy Hybrid Approaches with Other Decision-Making Variables

The first potential area of this dissertation aims to augment the developed hybrid fuzzy approaches by considering other decision variables and collecting more highway project data.

1) ***Increasing the sample size.***

   Due to the limited sample size of CM/GC highway projects in this dissertation, the results mainly referred to D-B-B and D-B highway projects. The reason is that state DOTs were asked to provide only completed CM/GC projects at the time of the data collection; however, CM/GC was still relatively new to many state DOTs at that time. In addition, other project delivery methods, such as public-private partnerships (P3), integrated project delivery (IPD), design-build-operate-maintain (DBOM), design-build-operate-transfer, and design-build-finance, should also be considered to enhance the applications of the hybrid fuzzy approaches in more project delivery decision-making scenarios.

Other project performance indicators, such as schedule growth, quality, maintenance costs, and sustainability problems, should also be considered in future work. The accommodation of these performance indicators can provide more insights regarding the advantages and disadvantages of the project delivery method candidates. However, adding these dependent variables may increase the complexity of the developed approaches as well as require more computational burdens. Increasing the number of dependent variables (which serve as cluster centers) leads to a larger number of the multi-dimensional spaces and more complex computations of distances within and between clusters.

2) ***Expanding the hybrid fuzzy approaches to other sectors.***

The hybrid fuzzy approaches can be applied to support the selection of project delivery methods in other horizontal project types, such as infrastructure and other transportation projects, and the vertical sector, including airport, commercial, and industrial projects. In the aviation sector, there is a need to deliver airport capital projects quickly from the planning phase to the design and construction phases. Due to the budget constraints, the number of deferred capital projects has been increased; as a result, many airports have progressively partnered with the private sector to transfer the responsibility for project delivery and performance. Currently, airports have applied various project delivery and financing methods to accelerate their major capital projects. Thus, there is a demand in the aviation sector to apply comprehensive decision-aid applications to project delivery method selection. Additional survey questionnaires, interviews, and case studies may be required to collect information regarding project complexity and delivery risk factors in those sectors.

Fuzzy Hybrid Approaches with Statistical and Machine Learning Techniques

The second potential area of this dissertation aims to improve the robustness of the developed hybrid fuzzy approaches by incorporating other statistical and machine learning techniques, such as Structural Equation Modeling (SEM), Monte Carlo simulation, and Bayesian belief networks.

3) ***Investigating underlying effects across project characteristics, delivery risks, project delivery methods, and cost performance.***

   The effects between input project characteristics, project complexity, and delivery risk factors on cost growth may impact the outcomes of the developed hybrid fuzzy approaches. To remove this limitation, SEM is a comprehensive statistical tool which helps model covariance between decision variables and assess the underlying interrelated relationships between their constructs (Mueller 2012). This technique is capable of producing the insights of observed and unobserved (or latent) variables simultaneously in terms of direct and indirect effects that exist between them (Raykov and Marcoulides 2012). For instance, in project delivery method selection, observed variables include project complexity and project cost growth while unobserved variables include project characteristics and delivery risks.

4) ***Validating and testing the hybrid fuzzy approaches.***

   Initially, the dataset contained 291 highway construction projects collected from 28 state DOTs. This study removed 37 projects identified with outliers in project cost data and delivery risks because fuzzy-based methods are very sensitive to outliers (Ross 2010). Analyzing reduced model predictions can support the verification, validation, and implementation of the developed hybrid fuzzy approaches. Monte Carlo simulation can help reduce the uncertainty in the prediction modeling process within the developed hybrid fuzzy approaches. Additionally, sensitivity analyses may also help identify the critical variables that affect the selection of project delivery methods. Future studies can implement Bayesian belief networks, which are established based on experts' opinions, to investigate the relationship between quantitative (project size and duration) and qualitative variables (project complexity and cost and schedule risk events).

Fuzzy Hybrid Approaches with Other Decision Scenarios in Construction

The third potential area of this dissertation aims to implement the developed hybrid fuzzy approaches to non-project-delivery-method decision-making scenarios in construction projects.

5) ***Expanding the hybrid fuzzy approaches to other decision-making scenarios.***

   This dissertation supports the selection of project delivery methods by accommodating a combination of quantitative and qualitative variables and investigating the causal relationship between them. The developed data-driven fuzzy approaches can also be applied in other scenarios of decision-making under uncertainty, such as asset management, financial budgeting and planning decisions, design and structural engineering decisions, and procurement decisions in international construction projects.

INTEGRATED REFERENCES

Al Khalil, M.I. (2002), "Selecting the appropriate project delivery method using AHP", *International Journal of Project Management*, 20(6), pp. 469-474

Al Nahyan, M., Hawas, Y., Raza, M., Aljassmi, H., Maraqa, M., Basheerudeen, B., and Mohammad, M. (2018). "A fuzzy-based decision support system for ranking the delivery methods of mega projects." *International Journal of Managing Projects in Business*, *11*(1), 122-143.

Ammar, M., Zayed, T., and Moselhi, O. (2013). "Fuzzy-Based Life-Cycle Cost Model for Decision Making under Subjectivity." *J. Constr. Eng. Manage.*, *139*(5), 556-563.

Anderberg, M. R. (2014). *Cluster analysis for applications: Probability and mathematical statistics: A series of monographs and textbooks,* Academic Press, New York.

Bakht, M., and El-Diraby, T. (2015). "Synthesis of Decision-Making Research in Construction." *J. Constr. Eng. Manage.*, *141*(9), 04015027.

Bayraktar, M., and Hastak, M. (2009). "Bayesian Belief Network Model for Decision Making in Highway Maintenance: Case Studies." *J. Constr. Eng. Manage.*, *135*(12), 1357-1369.

Bezdek, J. C. (1993). Fuzzy models—What are they, and why? *IEEE Trans. Fuzzy Syst.*, 1(1), 1–6.

Bezdek, J. C., Keller, J., Krisnapuram, R., and Pal, N. R. (1999). "Fuzzy models and algorithms for pattern recognition and image processing." *The handbooks of fuzzy sets series*, D. Dubois and H. Prade, eds., Springer, New York.

Bypaneni, S. (2017). *A BN-based Decision Framework for Selecting Project Delivery Methods in Highway Construction* (Ph.D.). University of Kansas.

Bypaneni, S., and Tran, D. (2018). "Empirical Identification and Evaluation of Risk in Highway Project Delivery Methods." *Journal of Management in Engineering*, *34*(3), 04018007.

Chan, A., Chan, D., and Yeung, J. (2009). "Overview of the Application of "Fuzzy Techniques" in Construction Management Research." *J. Constr. Eng. Manage.*, *135*(11), 1241-1252.

Chan, D., Chan, J., and Ma, T. (2014). Developing a fuzzy risk assessment model for guaranteed maximum price and target cost contracts in South Australia. *Facilities*, *32*(11/12), 624-646.

Chen, Q., Jin, Z., Xia, B., Wu, P., and Skitmore, M. (2016). "Time and Cost Performance of Design–Build Projects." *J. Constr. Eng. Manage.*, *142*(2), 04015074.

Chen, Y. Q., Liu, J. Y., Li, B. G., and Lin, B. S. (2011). "Project delivery system selection of construction projects in China." *Expert Syst. Appl.*, 38(5), 5456–5462

Cheng, M., Wu, Y., Wu, Y., and Ndure, S. (2019). "Fuzzy Bayesian schedule risk network for offshore wind turbine installation." *Ocean Engineering*, *188*, 106238.

Chiu, S. L. (1997). "Extracting fuzzy rules from data for function approximation and pattern classification." *Fuzzy information engineering: A guided tour of applications*, D. Dubois, H. Prade, and R. Yager, eds., Wiley, New York, 149–162.

Choi, J., Yun, S., Leite, F., and Mulva, S. (2019). "Team Integration and Owner Satisfaction: Comparing Integrated Project Delivery with Construction Management at Risk in Health Care Projects." *Journal of Management in Engineering*, *35*(1), 05018014.

Col Debella, D. M., and Ries, R. (2006). "Construction delivery systems: A comparative analysis of their performance within school districts." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2006)132:11(1131), 1131–1138.

Construction Industry Institute (CII). (2018). *Revisiting Project Delivery Performance*. McLean, VA: Charles Pankow Foundation.

Corona-Suárez, G., AbouRizk, S., 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*, *2014*, 1-18.

Creedy, G., Skitmore, M., and Wong, J. (2010). “Evaluation of risk factors leading to cost overrun in delivery of highway construction projects.” *J. Constr. Eng. Manage.*, 136(5), 528-537.

D’Urso, P. (2007). Fuzzy Clustering of Fuzzy Data. *Advances in Fuzzy Clustering and its Applications*. Eds. J. V. de Oliveira and W. Pedrycz. West Sussex, England: Wiley, 155 – 192.

Daneshvar, M. (2011). “Programmable trapezoidal and Gaussian membership function generator.” *J. Basic. Appl. Sci. Res.*, *1*(11), 2073-2079.

Das, S. (2013). “Pattern recognition using the fuzzy c-means technique.” *International Journal of Energy, Information and Communications*, *4*(1).

Diab, M., Varma, A., and Nassar, K. (2012). “Using risk assessment to improve highway construction project performance.” *48th ASC Annual Int. Conf. Proc.*, Associated Schools of Construction, Fort Collins, CO.

Douglas, A., Antoine, A., Schrilla, M., and Molenaar, K. (2016) “The Use and Performance of Alternative Contracting Methods on Small Highway Construction Projects.” *Procedia Engineering 145*, 908–915.

Duda, R. O., Hart, P. E., and Stork D. G. (2001). *Pattern Classification*. New York: Wiley

El Asmar, M., Hanna, A., and Loh, W. (2013). “Quantifying performance for the integrated project delivery system as compared to established delivery systems.” *J. Constr. Eng. Manage.* 139(11).

Elbarkouky, M., Fayek, A., Siraj, N., and Sadeghi, N. (2016). “Fuzzy Arithmetic Risk Analysis Approach to Determine Construction Project Contingency.” *J. Constr. Eng. Manage.*, *142*(12), 04016070.

Eleye-Datubo, A.G., Wall, A. and Wang, J. (2008). “Marine and offshore safety assessment by incorporative risk modeling in a fuzzy-Bayesian network of an induced mass assignment paradigm.” *Risk Analysis*, *28*(1), 95-112.

Elwood, E. (2014). “Fuzzy classification and fuzzy pattern recognition of seismic damage to concrete structures.” ⟨http://www/proquest.com/products-services/dissertations/⟩ (Dec. 31, 2018).

Elwood, E., and Corotis, R. (2015). “A Clustering Approach to Identification of Seismic Building Damage Patterns for Concrete Structures.” In *12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12*. Vancouver, Canada.

Elwood, E., and Corotis, R. (2015). “Application of Fuzzy Pattern Recognition of Seismic Damage to Concrete Structures.” *ASCE-ASME Journal oj Risk and Uncertainty In Engineering Systems, Part A: Civil Engineering*, *1*(4), 04015011.

Ersel, D., and İçen, D. (2014). “Fuzzy probability calculation with confidence intervals in Bayesian networks.” *Soft Computing*, *20*(2), 819-829.

Fayek, A. R., and A. Oduba. (2005). “Predicting industrial construction labor productivity using fuzzy expert systems.” *J. Constr. Eng. Manage*. 131 (8): 938–941

Ferreira, L., and Borenstein, D. (2012). “A fuzzy-Bayesian model for supplier selection.” *Expert Systems with Applications*, *39*(9), 7834-7844.

FHWA (2018), Tech Brief, Alternative Contracting Method Performance in U.S. Highway Construction, Federal Highway Administration, Washington, DC.

Garibaldi, J., Chen, C., and Razak, T. (2017). *Fuzzy Logic Toolkit for R* [Ebook]. CRAN. Retrieved from https://cran.r-project.org/web/packages/FuzzyR/FuzzyR.pdf

Goodrum, P., Uddin, M., and Faulkenberg, B. (2011). *A Case Study Analysis of the Kentucky Transportation Cabinet's Design/Build Pilot Projects*. Lexington, KY: Kentucky Transportation Center Research.

Gransberg, D.D., and J.S. Shane (2010). “Construction Manager-at-Risk Project Delivery for Highway Programs.” *NCHRP Synthesis 402*, ISBN: 978-9-14301-1, 2010, 127

Hale, D. R., Shrestha, P. P., Gibson, G. E., Jr., and Migliaccio, G. C. (2009). "Empirical comparison of design/build and design/bid/build project delivery methods." *J. Constr. Eng. Manage.*, 135(7), 579-587

Hastie, T., Tibshirani, R., and Friedman, J. (2009*). The elements of statistical learning: Data mining, inference, and prediction*, 2nd Ed., Springer, New York.

Hegazy, T. and Ayed, A. (1998). Neural network model for parametric cost estimation of highway projects. *J. Constr. Eng. Manage.*, 210 –218.

Hoppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999). *Fuzzy cluster analysis methods for classification, data analysis and image recognition*, Wiley, New York.

Hwang, H. (2004). “Identification of a Gaussian Fuzzy Classifier.” *International Journal of Control, Automation, And Systems*, *2*(1).

Ibbs, C. W., Kwak, Y. H., Ng, T., and Odabasi, A. M. (2003). “Project delivery systems and project change: Quantitative analysis.” *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2003)129:4(382), 382–387

İçen, D., and Ersel, D. (2019). “A new approach for probability calculation of fuzzy events in Bayesian Networks.” *International Journal of Approximate Reasoning*, *108*, 76-88.

Islam, M., and Nepal, M. (2016). “A Fuzzy-Bayesian Model for Risk Assessment in Power Plant Projects.” *Procedia Computer Science, 100*, 963-970.

Kassambara, A. (2017). “*Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis).*” <https://www.datanovia.com/en/wp-content/uploads/dn-tutorials/book-preview/clustering_en_preview.pdf> (Dec. 15, 2018)

Kawamura, H. (1993). “Fuzzy network for decision support systems.” *Fuzzy Sets and Systems*, *58*(1), 59-72.

Khanzadi, M., Nasirzadeh, F., Hassani, S., and Nejad Mohtashemi, N. (2016). “An integrated fuzzy multi-criteria group decision making approach for project delivery system selection.” *Scientia Iranica*, *23*(3), 802-814.

Khodakarami, V., and Abdi, A. (2014). “Project cost risk analysis: A Bayesian networks approach for modeling dependencies between cost items.” *International Journal of Project Management*, *32*(7), 1233-1245.

Kim, S. (2011). "Bayesian Model for Cost Estimation of Construction Projects." *Journal of the Korean Institute of Building Construction*, 10.5345/JKIC.2011.02.1.091, 91-99.

Klir, G. J., and Yuan, B. (1995). *Fuzzy sets and fuzzy logic theory and applications*, Prentice Hall, Upper Saddle River, NJ.

Knight, K., and Robinson Fayek, A. (2002). "Use of Fuzzy Logic for Predicting Design Cost Overruns on Building Projects." *J. Constr. Eng. Manage.*, *128*(6), 503-512

Knott, C., Hovell, L., and Karimian, N. (2013). *Type 1 Fuzzy Logic Toolkit* [Ebook]. CRAN. Retrieved from https://cran.r-project.org/web/packages/FuzzyToolkitUoN/FuzzyToolkitUoN.pdf

Konchar, M., and Sanvido, V. (1998). "Comparison of US project delivery systems." *J. Constr. Eng. Manag. ASCE* 124 (6), 435–444.

Kruse, R., Doring, C., and Lesot, M.-J. (2007). "Fundamentals of fuzzy clustering." *Advances in fuzzy clustering and its applications,* J. V. de Oliveira and W. Pedrycz, eds., Wiley, West Sussex, U.K., 3–30.

Kumar, D., Singh, J., Singh, O., and Seema. (2013). "A fuzzy logic based decision support system for evaluation of suppliers in supply chain management practices." *Mathematical and Computer Modelling*, *58*(11-12), 1679-1695.

Lam, K., So, A., Hu, T., Ng, T., Yuen, R., and Lo, S. et al. (2001). "An integration of the fuzzy reasoning technique and the fuzzy optimization method in construction project management decision-making." *Construction Management and Economics*, *19*(1), 63-76.

Lantz, B. (2015). *Machine Learning with R* (2nd ed., pp. 286-296). Birmingham, U.K.: Packt Publishing.

Li P.C., Chen G.H., Dai L.C., and Zhang L. (2012). "A fuzzy Bayesian network approach to improve the quantification of organizational influences in HRA frameworks." *Safety Science*, 50(7):1569–1583.

Li, J., Moselhi, O., and Alkass, S. (2006). “Forecasting Project Status by Using Fuzzy Logic.” *J. Constr. Eng. Manage.*, *132*(11), 1193-1202.

Li, T., Ng, S., and Skitmore, M. (2013). Evaluating stakeholder satisfaction during public participation in major infrastructure and construction projects: A fuzzy approach. *Automation in Construction*, *29*, 123-135.

Liu, K., Kuo, J., Yeh, K., Chen, C., Liang, H., and Sun, Y. (2013). “Using fuzzy logic to generate conditional probabilities in Bayesian belief networks: a case study of ecological assessment.” *International Journal of Environmental Science and Technology*, *12*(3), 871-884.

Love, P., Wang, X., Sing, C., and Tiong, R. (2013). “Determining the Probability of Project Cost Overruns.” *J. Constr. Eng. Manage.*, *139*(3), 321-330.

Luu, V., Kim, S., Tuan, N., and Ogunlana, S. (2009). "Quantifying schedule risk in construction projects using Bayesian belief networks." *International Journal of Project Management*, 10.1016/j.ijproman.2008.03.003,39-50.

Mafakheri, F., Dai, L., Slezak, D., and Nasiri, F. (2007). “Project delivery system selection under uncertainty multicriteria multilevel decision aid model.” *J. Manage. Eng.*, 4(200), 200–206.

Mahdi, I., and Alreshaid, K. (2005). “Decision Support system for selecting the proper delivery method using analytical hierarchy process (AHP).” *Int. J. Proj. Manage.*, 23, 564–572.10.1016/j.ijproman.2005.05.007

Mamdani, E. H., and Assilian, S. (1975). “An experiment in linguistic synthesis with a fuzzy logic controller.” *Int. J. Man-Machine Stud.*, 7(1), 1–13.

Martin, H., Lewis, T., Petersen, A., and Peters, E. (2017). “Cloudy with a Chance of Fuzzy: Building a Multicriteria Uncertainty Model for Construction Project Delivery Selection.” *Journal of Computing In Civil Engineering*, *31*(1), 04016046.

Marzouk, M., and Moselhi, O. (2004). Fuzzy Clustering Model for Estimating Haulers’ Travel Time. *J. Constr. Eng. Manage.*, *130*(6), 878-886.

Mathew, S., Tran, D., and Nguyen L.D. (2018) “Examining Impact of Risk on Cost Growth in Design-Build Projects Using Structural Equation Modeling,” *ASCE Construction Research Congress*, New Orleans, Louisiana, April 2 – 4.

Minchin, R., Li, X., Issa, R., and Vargas, G. (2013). “Comparison of Cost and Time Performance of Design-Build and Design-Bid-Build Delivery Systems in Florida.” *J. Constr. Eng. Manage.*, *139*(10), 04013007.

Molenaar, K., and Songer, A. (1998). “Model for Public Sector Design-Build Project Selection.” *J. Constr. Eng. Manage.*, *124*(6), 467-479.

Mostafavi, A. and Karamouz, M. (2010) “Selecting appropriate project delivery system: fuzzy approach with risk analysis.” *J. Constr. Eng. Manage.*, 136(8), 923–930.

Mostafavi, A., Abraham, D., Noureldin, S., Pankow, G., Novak, J., Walker, R., Hall, K., and George, B. (2013). “Risk-Based Protocol for Inspection of Transportation Construction Projects Undertaken by State Departments of Transportation.” *J. Constr. Eng. Manage.*, 139(8), 977–986.

Mueller, R. O. (2012). *Basic Principles of structural equation modeling: An introduction to LISREL and EQS*. New York. Springer Publications.

Nagalakshmi, G., and Jyothi, S. (2013). “A Survey on Pattern Recognition using Fuzzy Clustering Approaches.” *International Refereed Journal of Engineering and Science, 2*(5), 24-31.

Nguyen, L., Tran, D., and Chandrawinata, M. (2016). “Predicting Safety Risk of Working at Heights Using Bayesian Networks.” *J. Constr. Eng. Manage.*, *142*(9), 04016041.

Nguyen, P., Tran, D., and Lines, B. (2020). “Fuzzy Set Theory Approach to Classify Highway Project Characteristics for Delivery Selection.” *J. Constr. Eng. Manage.*, *146*(5), 04020044.

Nikou Goftar, V., El Asmar, M., and Bingham, E. (2014). “A meta-analysis of literature comparing project performance between design-build (DB) and design-bid-build (DBB) delivery systems.” *Construction Research Congress,* CRC Press, Atlanta

Niskanen, V. S. (2004). *Soft computing methods in human sciences*, Springer, New York.

Novák, V., Perfilieva, I., and Dvoøák, A. (2016). *Insight into Fuzzy Modeling*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Paek, J., Lee, Y., and Napier, T. (1992). “Selection of Design/Build Proposal Using Fuzzy-Logic System.” *J. Constr. Eng. Manage*., *118*(2), 303-317.

Pal, N. R., and Bezdek, J. C. (1995). “On cluster validity for the fuzzy c-means model.” *IEEE Trans. Fuzzy Syst.*, 3(3), 370–379.

Park, J., and Kwak, Y. H. (2017). “Design-bid-build (DBB) vs. design-build (DB) in the U.S. public transportation projects: The choice and consequences.” *Int. J. Project Management*, 35(3), 280–295

Pawan, P., and Lorterapong, P. (2016). "A Fuzzy-Based Integrated Framework for Assessing Time Contingency in Construction Projects." *J. Constr. Eng. Manage.*, *142*(3), 04015083

Piegat, A. (2001). *Fuzzy modeling and control*, Heidelberg, New York

Raykov, T., and Marcoulides, G. (2012). *A First Course in Structural Equation Modeling*, pp.1-52. Hoboken: Taylor and Francis.

Ren J., Jenkinson I., and Wang J. (2009). "An offshore risk analysis method using fuzzy Bayesian network." *Journal of Offshore Mechanics and Arctic Engineering*, 131(4), 1–12.

Říhová, E., and Makhalova, T. (2017). "On Evaluating of Fuzzy Clustering Results." In *The 11th International Days of Statistics and Economics*. Prague, Czech Republic.

Riza, L., Bergmeir, C., Herrera, F., and Benitez, J. (2014). "Learning from Data Using the R Package "frbs"." In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Beijing, China: IEEE.

Rojas, E. M., and Kell, I. (2008). "Comparative analysis of project delivery systems cost performance in pacific northwest public schools." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2008)134:6(387), 387–397

Ross, T. J. (2010). *Fuzzy logic with engineering applications*, 3rd Ed., Wiley, West Sussex, U.K.

Sargent, R.G. (2013). "Verification and validation of simulation models." *Journal of simulation*, 7(1), 12-24.

Scutari, M. (2009). "Learning Bayesian networks with the bnlearn R package." *arXiv preprint arXiv:0908.3817*.

Sedki, K., Delcroix, V., Lepoutre, F.X., Adam, E., Maquinghen-Godillon, A.P. and Ville, I. (2010). "Bayesian network model for decision problems." *Intelligent information systems*, 285-298.

Seo, S., Aramaki, T., Hwang, Y., and Hanaki, K. (2004). "Fuzzy Decision-Making Tool for Environmental Sustainable Buildings." *J. Constr. Eng. Manage.*, *130*(3), 415-423.

Setnes, M., Babuska, R., Kaymak, U., and van Nauta Lemke, H. (1998). "Similarity measures in fuzzy rule base simplification." *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), 28*(3), 376-386.

Shrestha, P., Migliaccio, G., O'Connor, J., and Gibson, G. (2007). "Benchmarking of Large Design–Build Highway Projects." *Transportation Research Record: Journal of the Transportation Research Board*, *1994*(1), 17-25.

Shrestha, P., O'Connor, J., and Gibson, G. (2012). "Performance Comparison of Large Design-Build and Design-Bid-Build Highway Projects." *J. Constr. Eng. Manage.*, *138*(1), 1-13.

Sullivan, J., Asmar, M. E., Chalhoub, J., and Obeid, H. (2017). "Two decades of performance comparisons for design-build, construction manager at risk, and design-bid-build: Quantitative analysis of the state of knowledge on project cost, schedule, and quality." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0001282, 04017009

Sun, D., Jia, Y., Yang, Y., Li, H., and Zhao, L. (2018). "Fuzzy-Bayesian-network-based Safety Risk Analysis in Railway Passenger Transport." *Periodica Polytechnica Transportation Engineering*, *46*(3), 135-141.

Sun, J., Liu, B., Chu, Z., Chen, L. and Li, X. (2018). "Tunnel collapse risk assessment based on multistate fuzzy Bayesian networks." *Quality and Reliability Engineering International,* 34(8), 1646-1662.

Thanathornwong, B. (2018). “Bayesian-Based Decision Support System for Assessing the Needs for Orthodontic Treatment.” *Healthcare Informatics Research*, *24*(1), 22.

Touran, A., Gransberg, D., Molenaar, K., and Ghavamifar, K. (2011). “Selection of Project Delivery Method in Transit: Drivers and Objectives.” *Journal of Management in Engineering*, 27(1), 21-27.

Touran, A., Gransberg, D.D., Molenaar, K.R., Ghavamifar, K., Mason, D.J., and Fithian, L.A. (2009a). “A guidebook for the evaluation of project delivery methods.” *Rep. No. TCRP Report 131*, Transportation Research Board.

Touran, A., Molenaar, K., Gransberg, D., and Ghavamifar, K. (2009b). “Decision Support System for Selection of Project Delivery Method in Transit.” *Transportation Research Record: Journal of the Transportation Research Board*, 2111, 148-157.

Tran, D. Q., Molenaar, K. R., and Alarcon, L. F. (2016) “A Hybrid Cross-Impact Approach to Predicting Cost and Variance of Project Delivery Decisions for Highways,” *ASCE Journal of Infrastructure Systems*, 22(1), 04015017-04015017-12

Tran, D., and Molenaar, K. (2015). Risk-Based Project Delivery Selection Model for Highway Design and Construction. *J. Constr. Eng. Manage.*, *141*(12), 04015041.

Tran, D., Diraviam, G., and Minchin, R. (2018). “Performance of Highway Design-Bid-Build and Design-Build Projects by Work Types.” *J. Constr. Eng. Manage.*, *144*(2), 04017112.

Tran, D., Harper C.M., Molenaar, K. R.; Haddad, N.F., and Scholfield, M.M. (2013). “A Project Delivery Selection Matrix for Highway Design and Construction” *Transportation Research Record: Journal of the Transportation Research Board, National Academies,* 2347, 3-10

Tran, V., Yang, B., Oh, M., and Tan, A. (2009). Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. *Expert Systems With Applications*, *36*(2), 1840-1849.

U.S. DOT, Federal Highway Administration. (2006). “Design-build effectiveness study.” ⟨http://www.fhwa.dot.gov/reports/designbuild /designbuild.pdf⟩ (Feb. 28, 2019).

Ung, S. (2018). “Development of a weighted probabilistic risk assessment method for offshore engineering systems using fuzzy rule-based Bayesian reasoning approach”. *Ocean Engineering*, *147*, 268-276.

Uusitalo, L. (2007). “Advantages and challenges of Bayesian networks in environmental modelling.” *Ecological modelling*, 203(3-4), 312-318.

Viattchenin, D., Tati, R., and Damaratski, A. (2013). “Designing Gaussian Membership Functions for Fuzzy Classifier Generated by Heuristic Possibilistic Clustering.” *J.I.O.S.*, *37*(2).

Viertl, R. (1987). Is it necessary to develop fuzzy Bayesian inference, In R. Viertl (Ed.), *Probability and Bayesian Statistics* (pp. 471–475). New York: Plenum Press

Viertl, R., and Sunanta, O. (2013). “Fuzzy Bayesian inference.” *METRON*, *71*(3), 207-216.

Weber, P., Medina-Oliva, G., Simon, C., and Iung, B. (2012). “Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas.” *Engineering Applications of Artificial Intelligence*, *25*(4), 671-682.

WSDOT (2016). *Project Delivery Method Selection Guidance*. Washington State Department of Transportation, Construction Division.

Wu, K., and Yang, M. (2005). “A cluster validity index for fuzzy clustering.” *Pattern Recognition Letters*, *26*(9), 1275-1291. doi: 10.1016/j.patrec.2004.11.022

Wu, X., Wu, B., Sun, J., and Fu, H. (2010). "Unsupervised Possibilistic Fuzzy Clustering." *Journal of Information and Computational Science*, *7*(5), 1075-1080.

Yang J., Liu J., Wang J., Sii H., and Wang H. (2006). "Belief rule-base inference methodology using the evidential reasoning Approach-RIMER." *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *36*(2), 266-285.

Yazdi, M., and Kabir, S. (2017). "A fuzzy Bayesian network approach for risk analysis in process industries." *Process Safety and Environmental Protection*, *111*, 507-519.

Ye, Z., Sadeghian, A., and Wu, B. (2006). Mechanical fault diagnostics for induction motor with variable speed drives using Adaptive Neuro-fuzzy Inference System. *Electric Power Systems Research*, *76*(9-10), 742-752.

Zadeh, L. A. (1965). "Fuzzy sets." *Inf. Control*, 8(3), 338–353.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.*, 1(1), 3–28.

Zarei, E., Khakzad, N., Cozzani, V., and Reniers, G. (2019). "Safety analysis of process systems using Fuzzy Bayesian Network (FBN)." *Journal of Loss Prevention in the Process Industries*, *57*, 7-16.

Zarikas, V., Papageorgiou, E., and Regner, P. (2014). "Bayesian network construction using a fuzzy rule based approach for medical decision support." *Expert Systems*, *32*(3), 344-369.

Zeng, Y., and Starzyk, J. (2018). "Statistical approach to clustering in pattern recognition." In *Proceedings of the 33rd Southeastern Symposium on System Theory*. Athens, OH, USA: IEEE.

Zhang, L., Wu, X., Qin, Y., Skibniewski, M.J. and Liu, W. (2016). “Towards a fuzzy Bayesian network based approach for safety risk analysis of tunnel-induced pipeline damage.” *Risk Analysis*, 36(2), 278-301.

Zoullouti, B., Amghar, M., and Sbiti, N. (2017). “Risk Analysis of Operating Room Using the Fuzzy Bayesian Network Model.” *International Journal of Engineering*, *30*(1).

APPENDIX A – FUZZY CLASSIFICATION

A.1. Exploratory Factor Analysis Result

Table A.1. Exploratory Factor Analysis of 31 Critical Delivery Risks (Bypaneni 2017)

Risk Factors	Loading	% of Variance	Cumulative (%)
Risk Factor 1: Complexity Risk		**24.99**	**24.99**
Project complexity	0.71		
Uncertainty in geotechnical investigation	0.59		
Legal challenges and changes in law	0.58		
Intergovernmental agreements and jurisdiction	0.57		
Difficulty in obtaining other agency	0.49		
Risk 2: Quality Risk		**10.91**	**35.90**
Construction QC/QA process	0.93		
Design Quality Assurance	0.88		
Risk 3: Constructability Risk		**10.15**	**46.05**
Delays in procuring critical materials, labor, and specialized equipment	0.85		
Significant increase in material, labor and equipment cost	0.76		
Risk 4: Construction Risk		**7.53**	**53.59**
Work zone traffic control	0.86		
Construction sequencing/staging/phasing	0.83		
Risk 5: Utility and ROW Risk		**6.69**	**60.28**
Unexpected utility encounter	0.76		
Delays in completing utility agreements	0.71		
Delays in right-of-way (ROW) process	0.63		
Risk 6: Management Risk		**5.79**	**66.07**
Staff experience/availability	0.80		
Project and program management issues	0.76		
Conformance with regulations/guidelines/design criteria	0.72		
Risk 7: Environmental Risk		**5.06**	**71.13**
Challenges to obtain appropriate environmental documentation	0.61		
Environmental impacts	0.48		

A.2. Determination of Number of Clusters

Selection of an optimal number of clusters is subjective and depends on methods of similarity measurement and clustering parameters. Table 2 provides results associated with four utilized methods of identifying the most optimal number of meaningful clusters. As a result of the four methods, the range of the potential numbers of clusters was from two to ten clusters, and the most selected optimal number of clusters was seven. Accordingly, seven clusters were pre-defined as the input for the number of cluster centers to the FCM algorithm.

- Elbow Method
    - Chooses a number of clusters so that adding another cluster doesn't improve much better the total WSS.

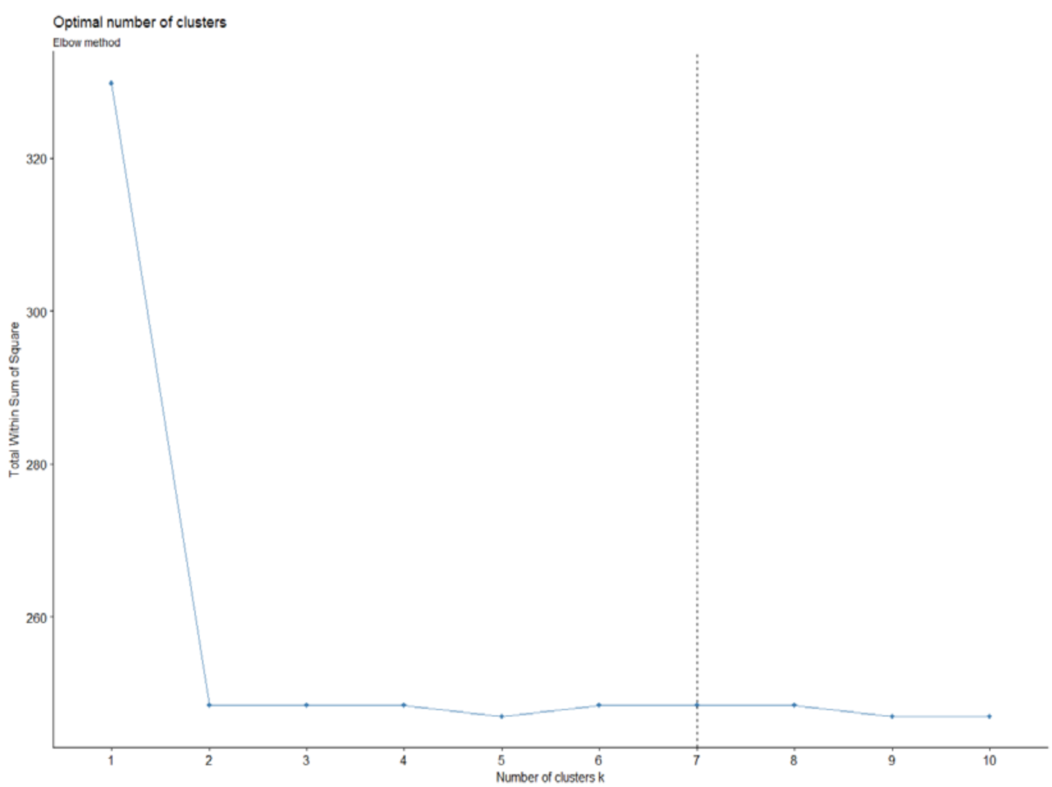

Figure A.2. Elbow method for fuzzy cluster analysis

- <u>Silhouette Method</u>
    - o Determines how well each object lies within its cluster. A high average silhouette width indicates a good clustering.

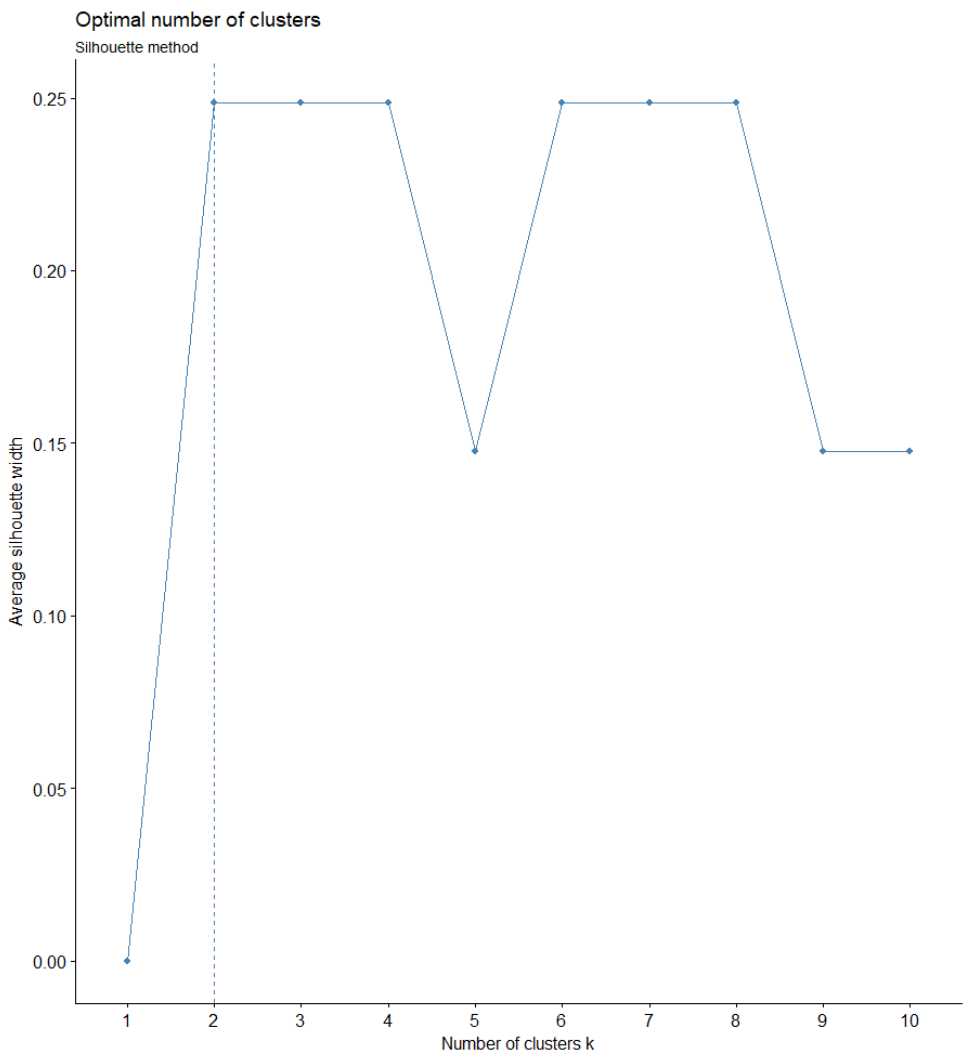

Figure A.3. Silhouette method for fuzzy cluster analysis

- Gap Statistics Method
    - Compares the total within intra-cluster variation for different values of k with their expected values under null reference distribution of the data. The estimate of the optimal clusters will be value that maximize the gap statistic (nboot = 50).

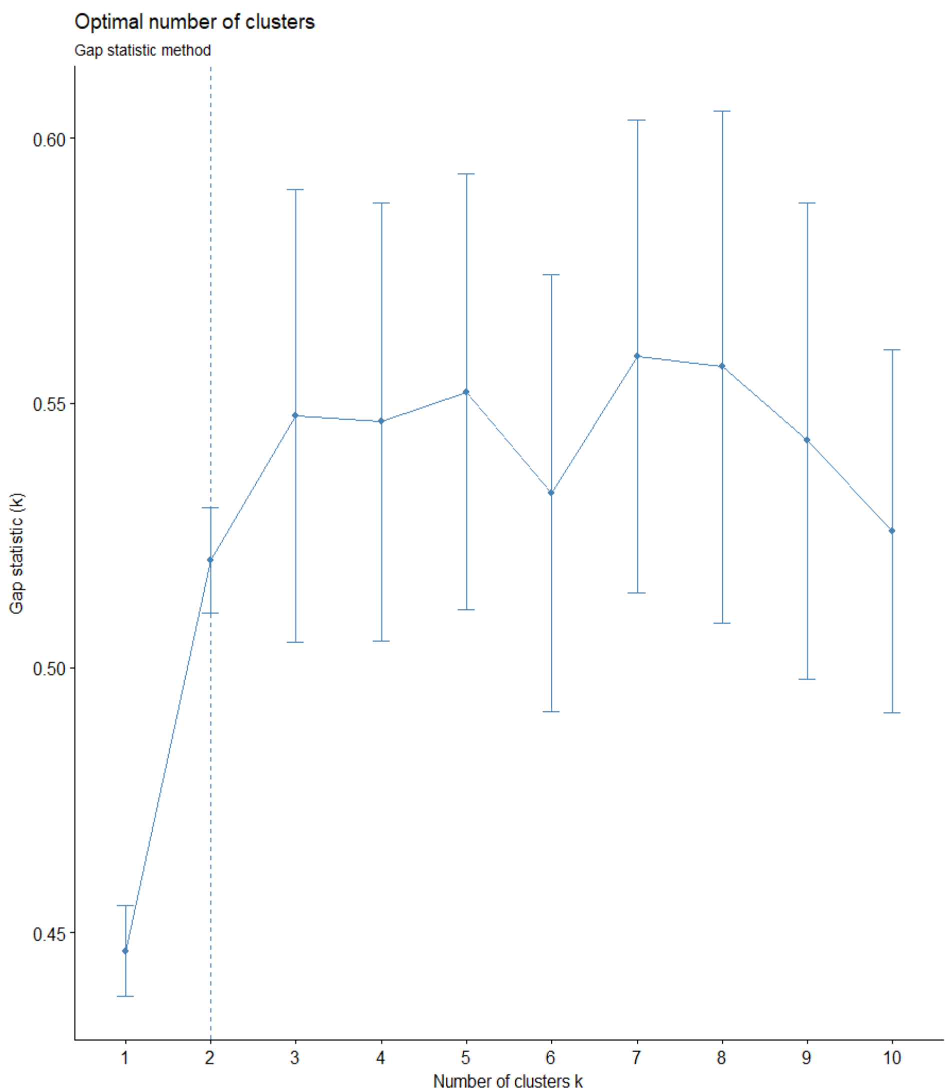

Figure A.4. Gap statistic method for fuzzy cluster analysis

- NbClust Method
  - Using 27 clustering indices to determine an appropriate number of clusters.

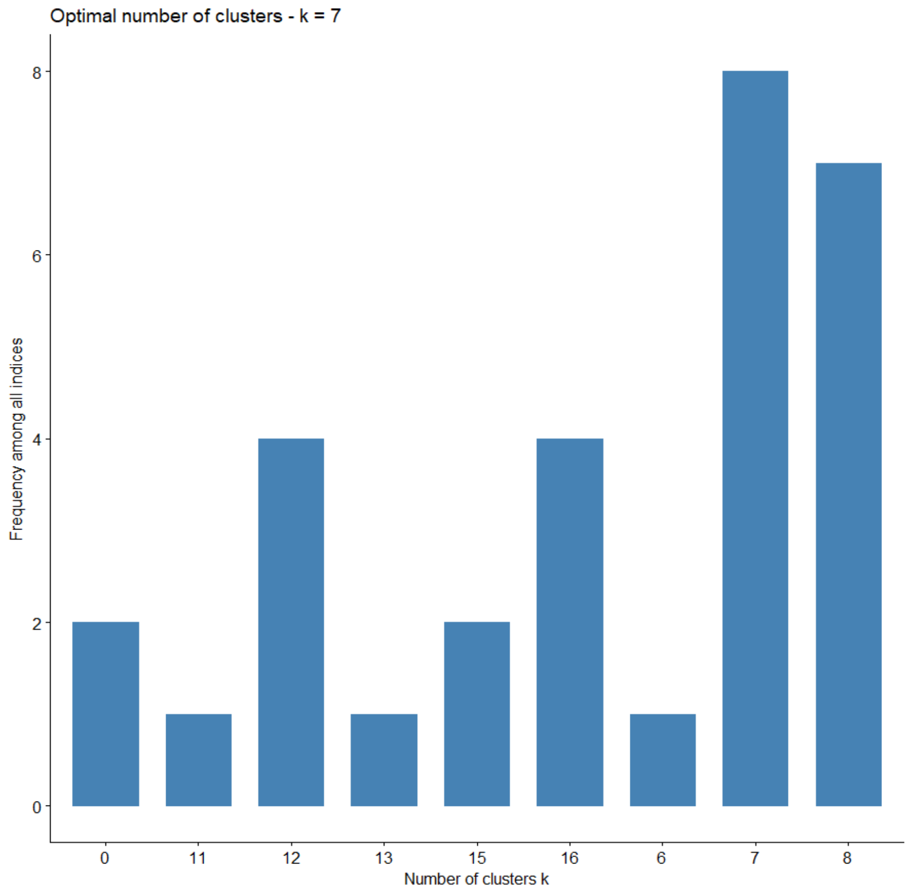

Figure A.5. NbClust method for cluster analysis

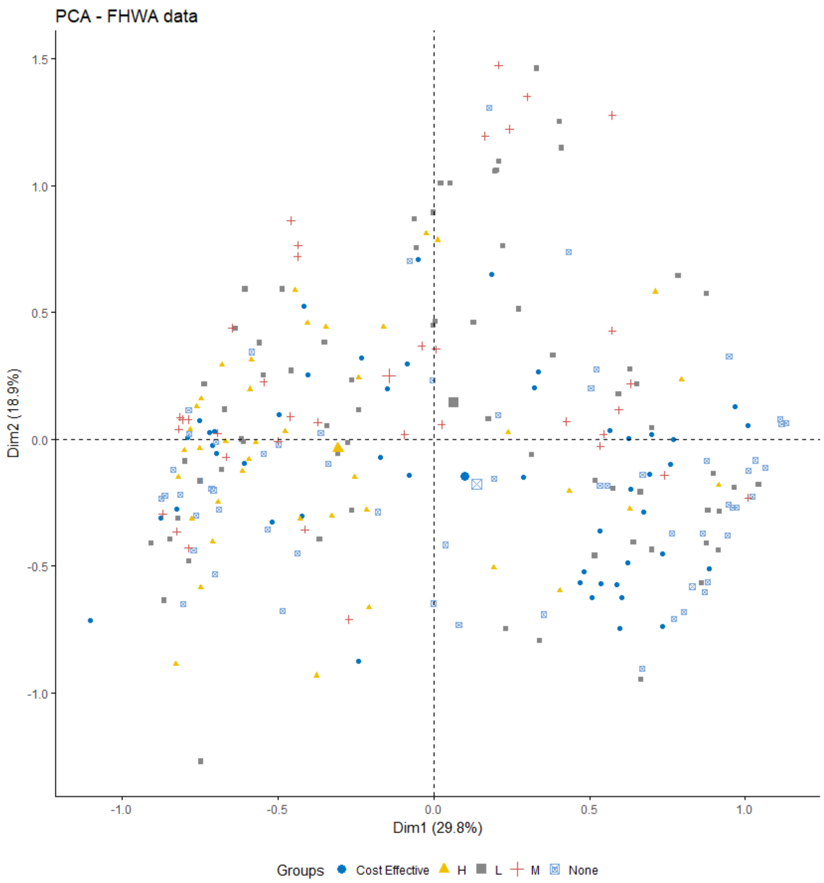

Figure A.6. Distribution of the Five Cost Growth Groups within the Collected Dataset

A.3. Fuzzy C-means Cluster Analysis Algorithm

- <u>Coding Procedure in R</u>
  - R package - Probabilistic and Possibilistic Cluster Analysis (ppclust): This method is the most advanced fuzzy cluster analysis up-to-date can optimize the issues of outliers and minimize overlapping clustering groups.
  - Cluster centers = 7, fuzziness (m) = 1.65, nstart = 5

```r
res.ppclust4 <- upfc(x = dat.std, centers = 7, m = 1.65, nstart = 5) # m = 1.536 or 1.53 or 1.62, nstart = 5
res.ppclust4$csize                # Number of objects in the clusters
res.ppclust4$u                    # Fuzzy Membership Matrix
res.ppclust4$v0                   # Initial Cluster Prototypes
res.ppclust4$v                    # Final Cluster Prototypes
res.ppclust4$d                    # Distances of objects to the final cluster prototypes
res.ppclust4$cluster              # cluster labels found by defuzzying the fuzzy membership degrees of the objects
res.ppclust4$sumsqrs$between.ss   # Between-cluster sum of squares
res.ppclust4$sumsqrs$within.ss    # Within-cluster sum of squares for each cluster
res.ppclust4$sumsqrs$tot.within.ss # Total within-cluster sum of squares
res.ppclust4$sumsqrs$tot.ss       # Total sum of squares
```

Figure A.7. Coding Demonstration of Fuzzy Cluster Analysis in R Programming

Table A.2. Number of objects in the clusters

Cluster 3	Cluster 4	Cluster 2	Cluster 7	Cluster 5	Cluster 6	Cluster 1
39	23	34	33	34	39	52

- Between-cluster sum of squares: 144.54

Table A.3. Within-cluster sum of squares for each cluster

Cluster 3	Cluster 4	Cluster 2	Cluster 7	Cluster 5	Cluster 6	Cluster 1
12.18	9.77	21.62	26.72	32.74	15.26	31.26

- Total within-cluster sum of squares: 149.53

  (between_SS / total_SS = 50.00%)

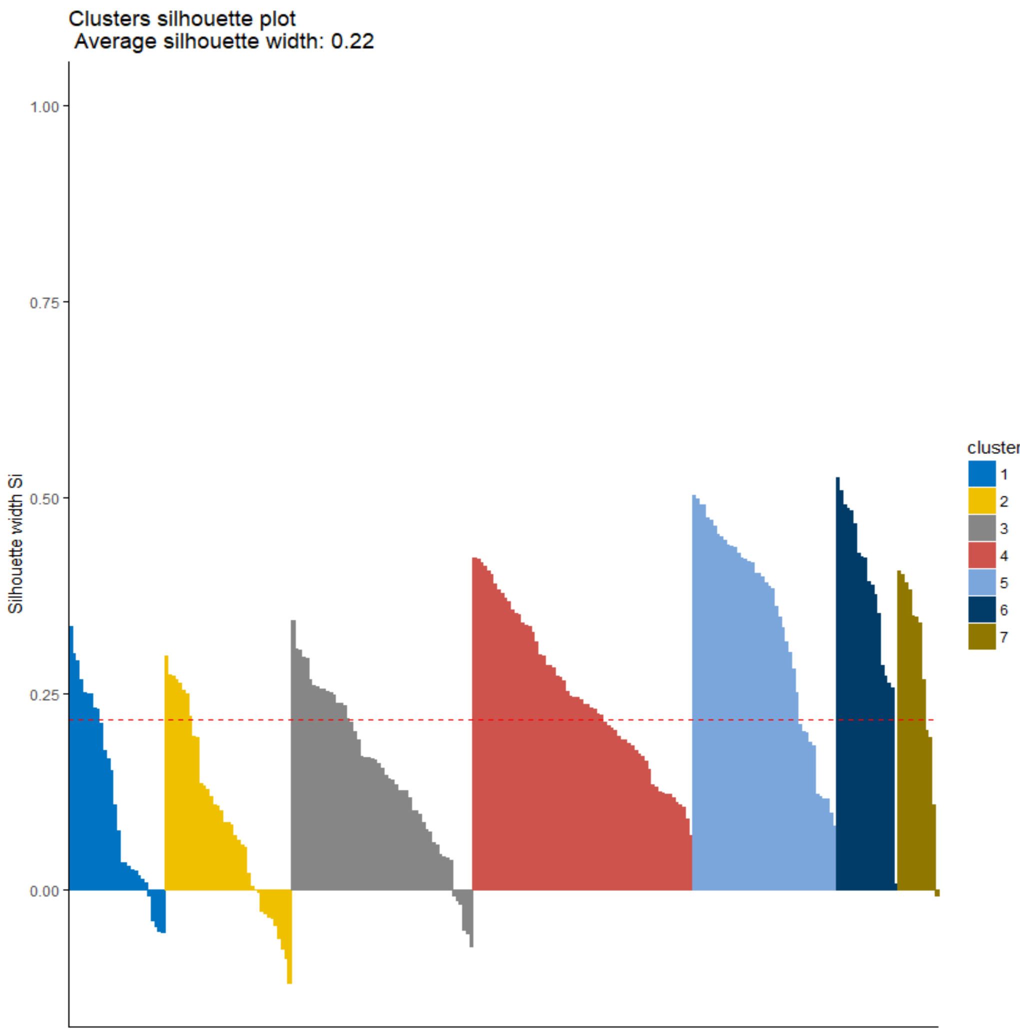

Figure A.8. Silhouette Plot of the Seven Identified Clusters

Table A.4 shows the cluster centers of 17 variables after hardening for interpretation. All of the data are hardened in terms of their original data type. The cluster centers of cost growths are then used to claim the cost performance patterns for subsequent analyses. Tables A.5 and A.6 demonstrate distributions of project delivery methods in established clusters with a cut-off point of 30%. Tables A.7 and A.8 demonstrate distributions of cost growth patterns in clusters with a cut-off point of 20%.

Table A.4. Terminal Cluster Center and Spread Values of the FHWA data set

Variables	Cluster 3	Cluster 4	Cluster 2	Cluster 7	Cluster 5	Cluster 6	Cluster 1
No. of Projects	39 (15.3%)	23 (9.1%)	34 (13.4%)	33 (13.0%)	34 (13.4%)	39 (15.3%)	52 (20.5%)
Fac-Road (%)	**67.97**	**62.98**	28.13	**41.45**	**65.52**	11.85	**59.14**
Fac-Bridge (%)	14.01	16.55	**55.25**	**30.55**	9.14	**83.52**	23.10
Fac-Drainage (%)	9.14	9.62	4.55	10.64	3.91	1.78	6.62
Fac-ITS (%)	2.56	3.03	2.68	6.37	1.57	0.32	2.03
Fac-Other (%)	6.36	7.85	9.41	11.02	19.89	2.54	9.13
Pro-New (%)	**89.02**	**86.17**	**81.66**	**67.29**	5.24	2.05	4.40
Pro-Reno (%)	7.44	9.32	10.20	21.93	8.52	**95.94**	**91.32**
Pro-Other (%)	3.54	4.51	8.13	10.78	**86.24**	2.00	4.29
Complexity	1.22	1.29	1.91	1.37	2.37	1.82	1.67
RF1	2.54	2.65	2.26	3.22	2.06	2.02	2.40
RF2	2.33	2.44	2.14	3.09	2.05	1.97	2.31
RF3	2.31	2.41	2.04	3.02	2.10	1.97	2.32
RF4	2.52	2.64	2.44	3.48	2.60	2.20	3.12
RF5	2.48	2.60	2.07	3.20	1.81	1.93	2.22
RF6	2.11	2.18	2.02	2.66	1.94	1.98	2.15
RF7	2.68	2.80	2.54	3.39	2.05	2.09	2.36
Cost Growth (%)	15	5	3	0	4	1	2
Spread (Fuzziness)	0.16	0.12	0.14	0.11	0.18	0.25	0.21

Table A.5. Distribution of Project Delivery Methods in Clusters (Count)

Cluster	DBB	DB	CMGC	Total	Dominance
3	24	12	3	39	DBB
4	7	13	3	23	DB
2	10	22	2	34	DB
7	9	17	7	33	DB
5	12	19	3	34	DB
6	18	17	4	39	DBB and DB
1	29	14	9	52	DBB
Total	109	114	31	254	-
Dominance	C1, C6, and C7	C3 and C5	C4 and C7	-	-

Table A.6. Distribution of Project Delivery Methods in Clusters (Cut-off point = 30%)

Cluster	DBB	DB	CMGC	Total	Dominance
3	**62%**	31%	8%	100%	DBB
4	30%	**57%**	13%	100%	DB
2	29%	**65%**	6%	100%	DB
7	27%	**52%**	21%	100%	DB
5	35%	**56%**	9%	100%	DB
6	**46%**	**44%**	10%	100%	DBB and DB
1	**56%**	27%	17%	100%	DBB

Table A.7. Distribution of Cost Growth Patterns in Clusters (Count)

Cluster	Cost Eff.	None	Low	Medium	High	Total	Dominance
3	5	9	5	**11**	**9**	39	Medium - High
4	**5**	**5**	**6**	2	5	23	CE - Low
2	7	6	**10**	4	7	34	Low
7	4	**8**	**9**	4	8	33	None - Low
5	4	3	**18**	6	3	34	Low
6	5	22	10	1	1	39	None
1	**18**	**7**	**12**	8	7	52	CE
Total	48	60	70	36	40	254	-
Dominance	C7	C6	C5	C1	C1, C3, and C7	-	-

Table A.8. Distribution of Cost Growth Patterns in Clusters (Cut-off point = 20%)

Cluster	Cost Eff.	None	Low	Medium	High	Total	Dominance
3	13%	23%	13%	**28%**	**23%**	100%	Medium - High
4	**22%**	**22%**	**26%**	9%	22%	100%	CE - Low
2	21%	18%	**29%**	12%	21%	100%	Low
7	12%	**24%**	**27%**	12%	24%	100%	None - Low
5	12%	9%	**53%**	18%	9%	100%	Low
6	13%	**56%**	26%	3%	3%	100%	None
1	**35%**	13%	23%	15%	13%	100%	CE

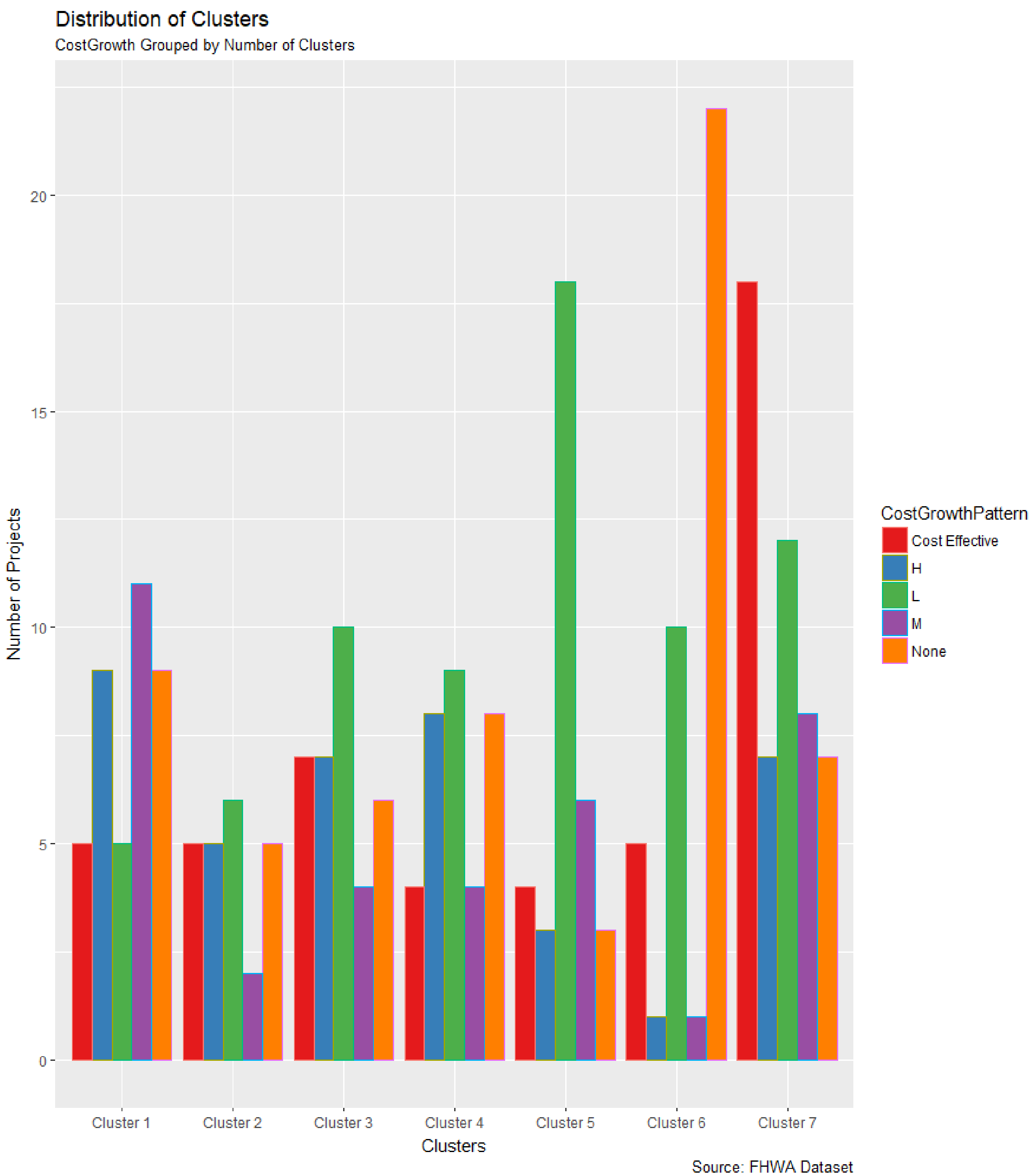

Figure A.9. Distributions of Five Groups of Cost Growth within Seven Clusters

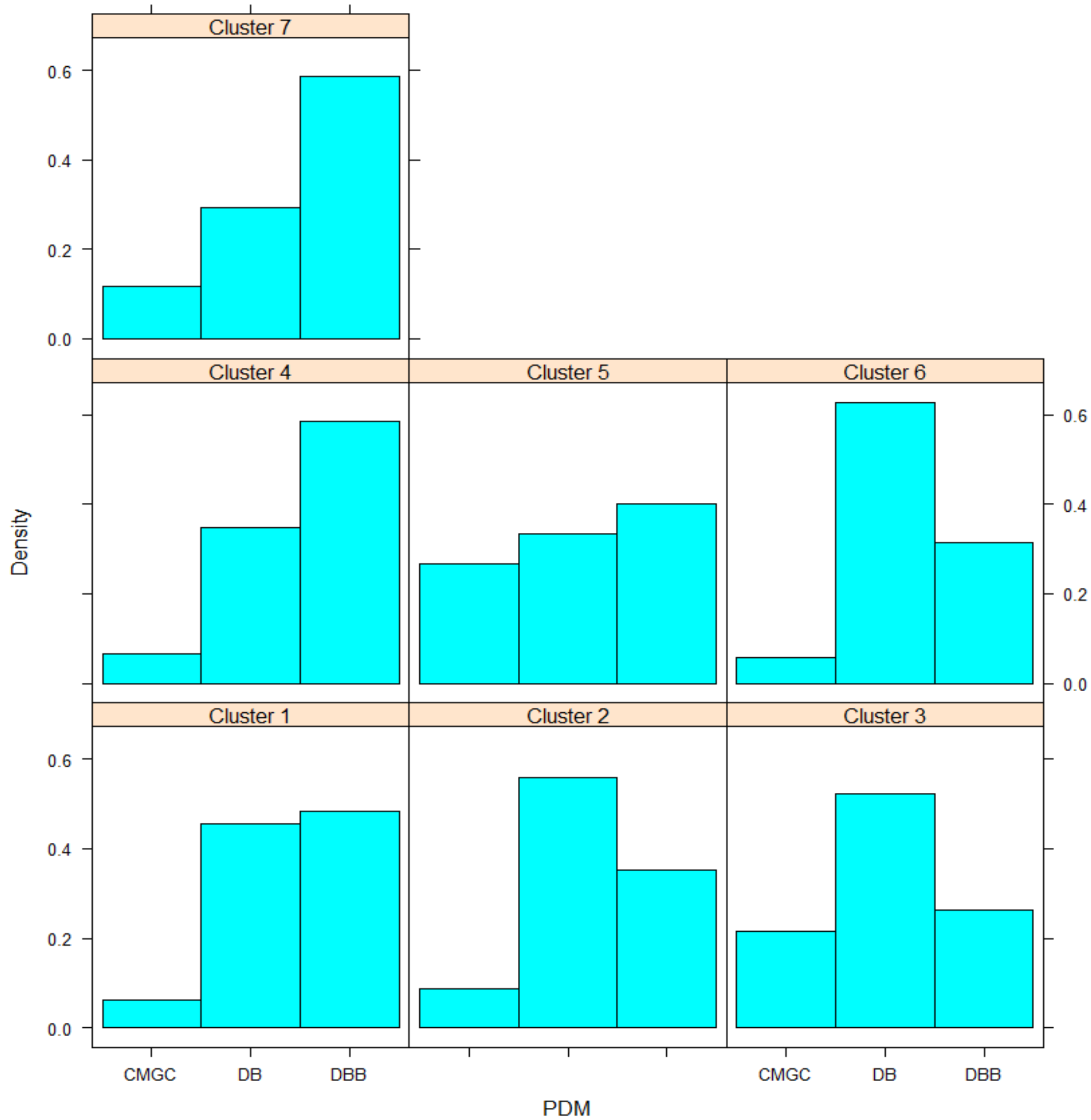

Figure A.10. Distributions of Project Delivery Methods within Seven Clusters

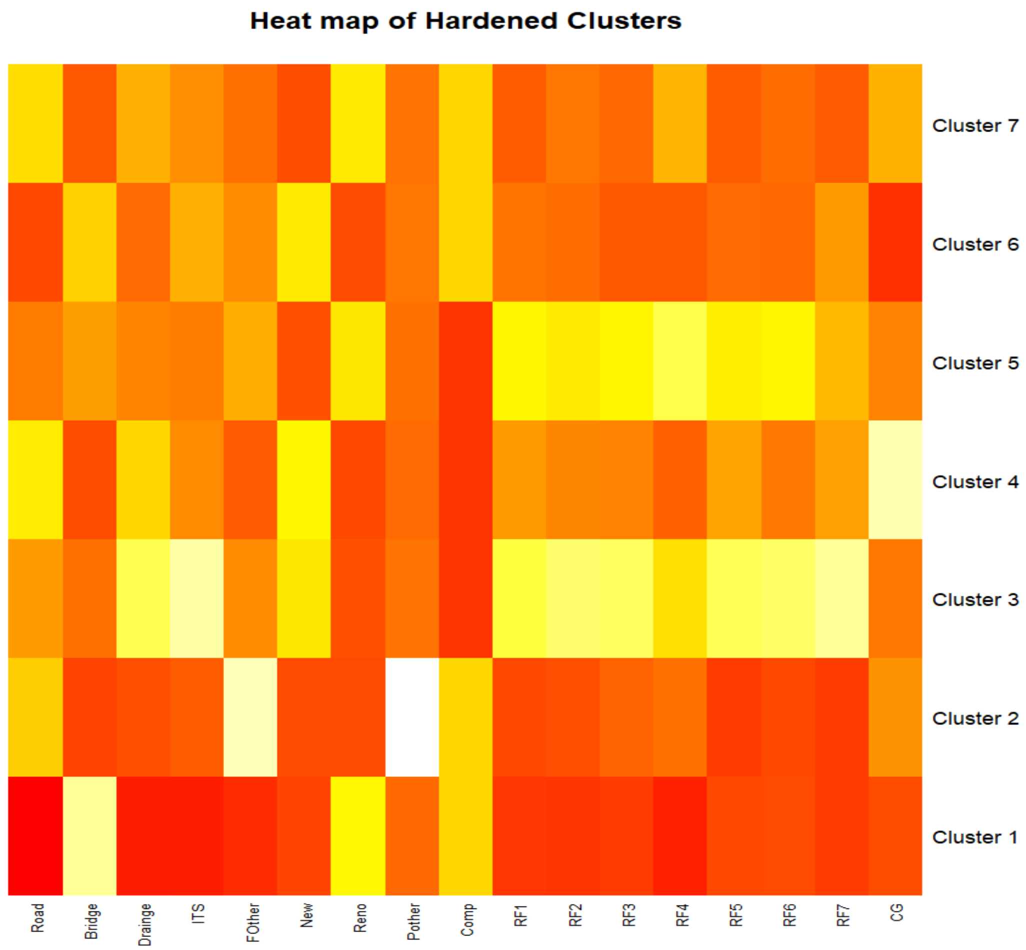

Figure A.11. Heat Map of 17 Variables within Seven Clusters

Table A.9. 17 Selected Variables.

Variable	Data Type	Mean	Min	Max
Facility Type - Road	Continuous	45.56	0	100
Facility Type - Bridge	Continuous	33.58	0	100
Facility Type - Drainage	Continuous	6.71	0	90
Facility Type – Intelligent Transportation Systems (ITS)	Continuous	3.08	0	100
Facility Type - Other	Continuous	11.10	0	100
Project Type – New	Continuous	45.47	0	100
Project Type – Reconstruction	Continuous	39.46	0	100
Project Type – Other	Continuous	15.07	0	100
Project Complexity	Ordinal	2	1	3
Risk Factor 1 - Complexity	Continuous	2.47	1	5.27
Risk Factor 2 - Quality	Continuous	2.36	1	6
Risk Factor 3 - Constructability	Continuous	2.33	1	5
Risk Factor 4 - Construction	Continuous	2.78	1	6
Risk Factor 5 - Utility and ROW	Continuous	2.37	1	6
Risk Factor 6 - Management	Continuous	2.16	1	5
Risk Factor 7 - Environmental	Continuous	2.57	1	6
Project Delivery Method	Nominal	-	-	-
Cost Performance	Continuous	0.03	-0.12	0.24

APPENDIX B – FUZZY PATTERN RECOGNITION

B.1. Fuzzy Membership Functions of All Variables

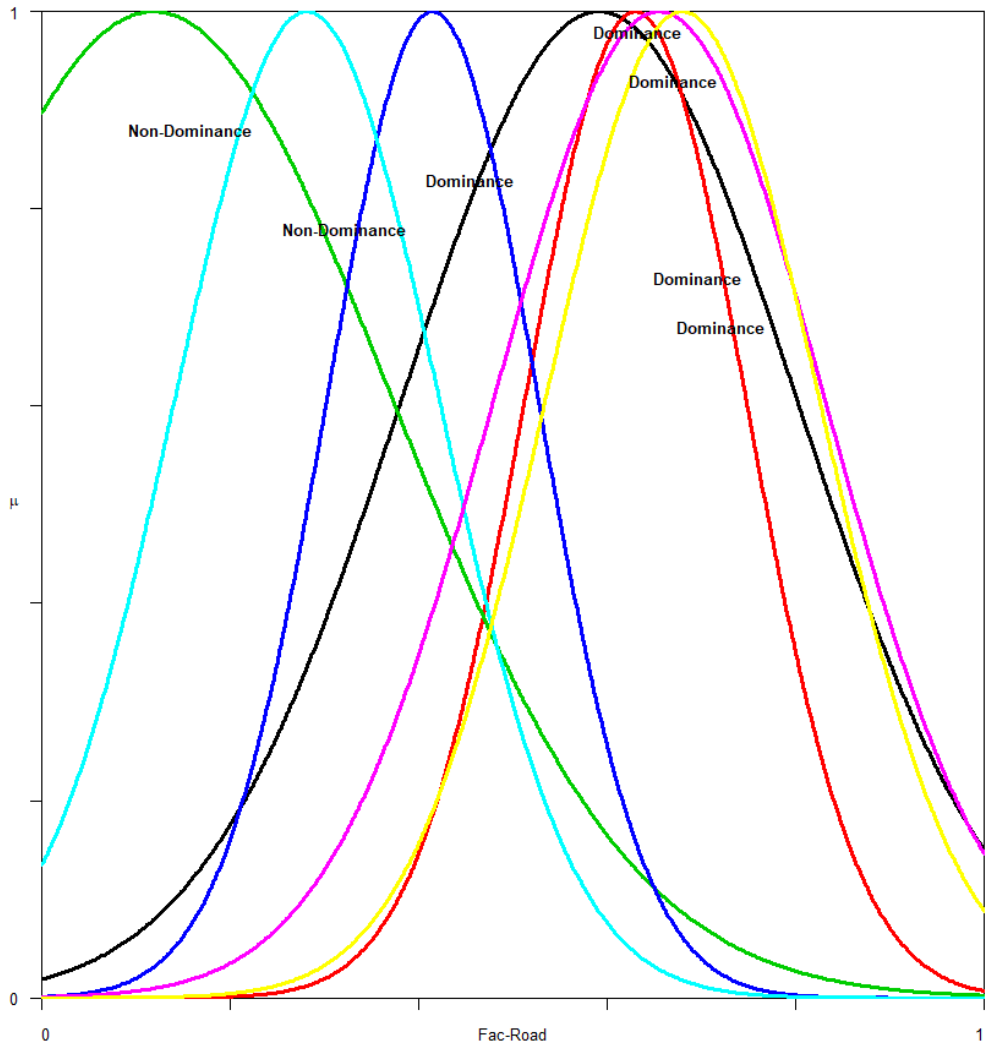

Figure B.1. Membership Function of Fac-Road

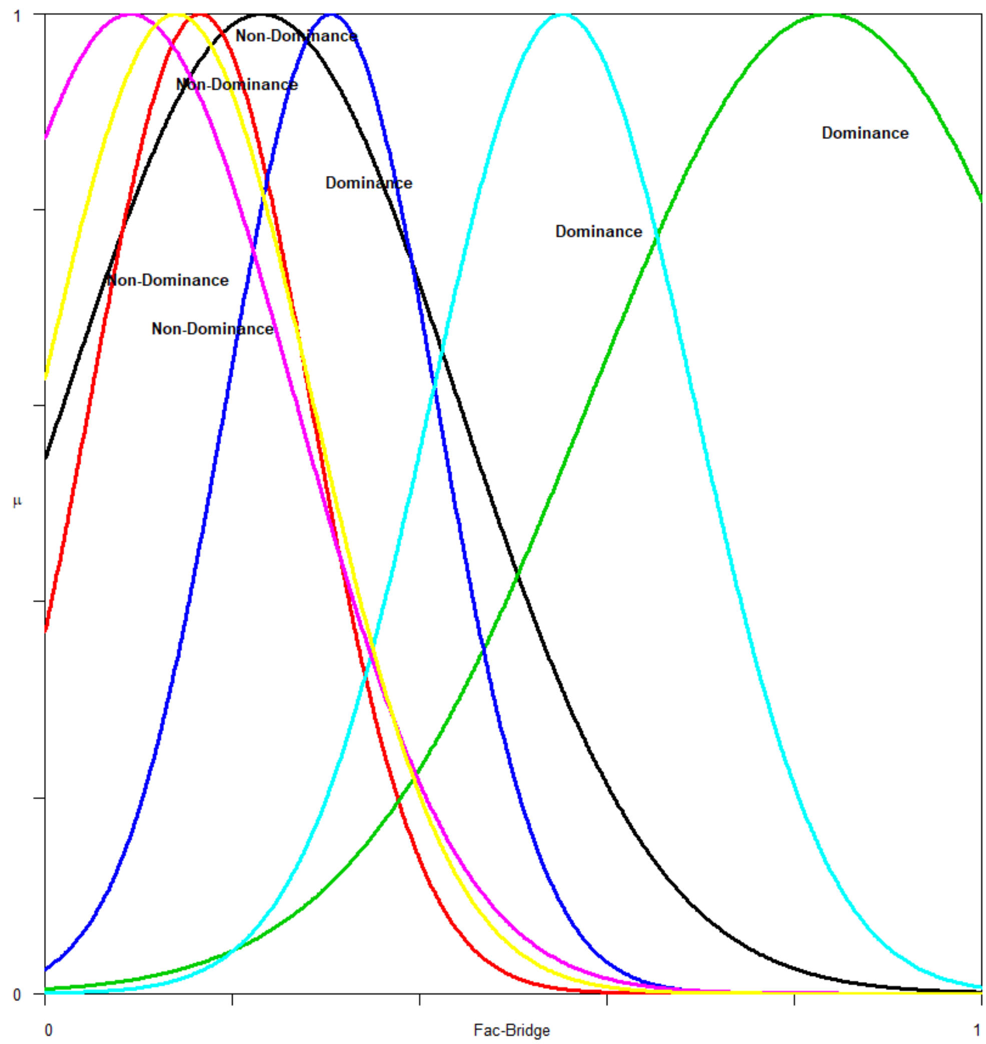

Figure B.2. Membership Function of Fac-Bridge

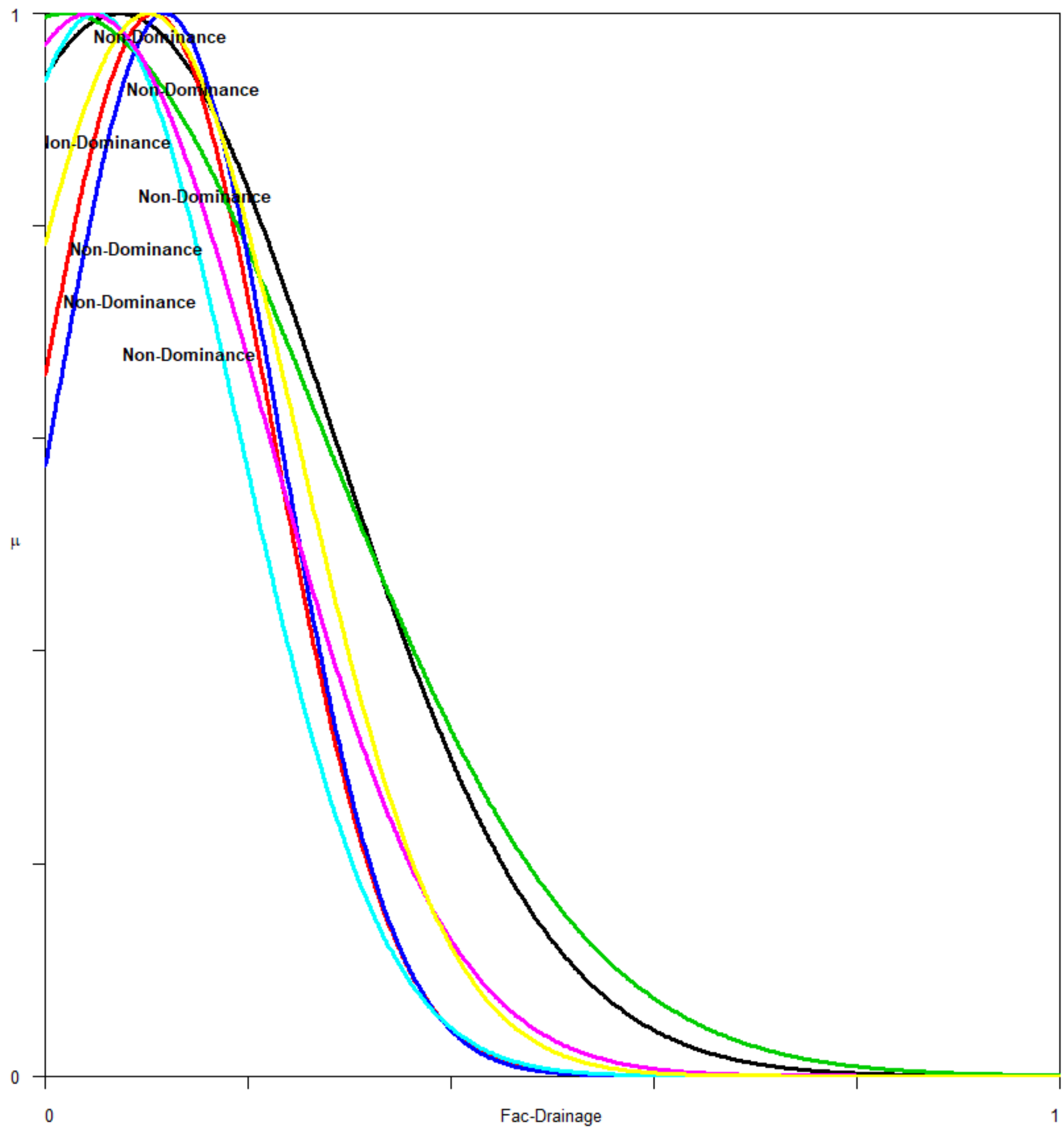

Figure B.3. Membership Function of Fac-Drainage

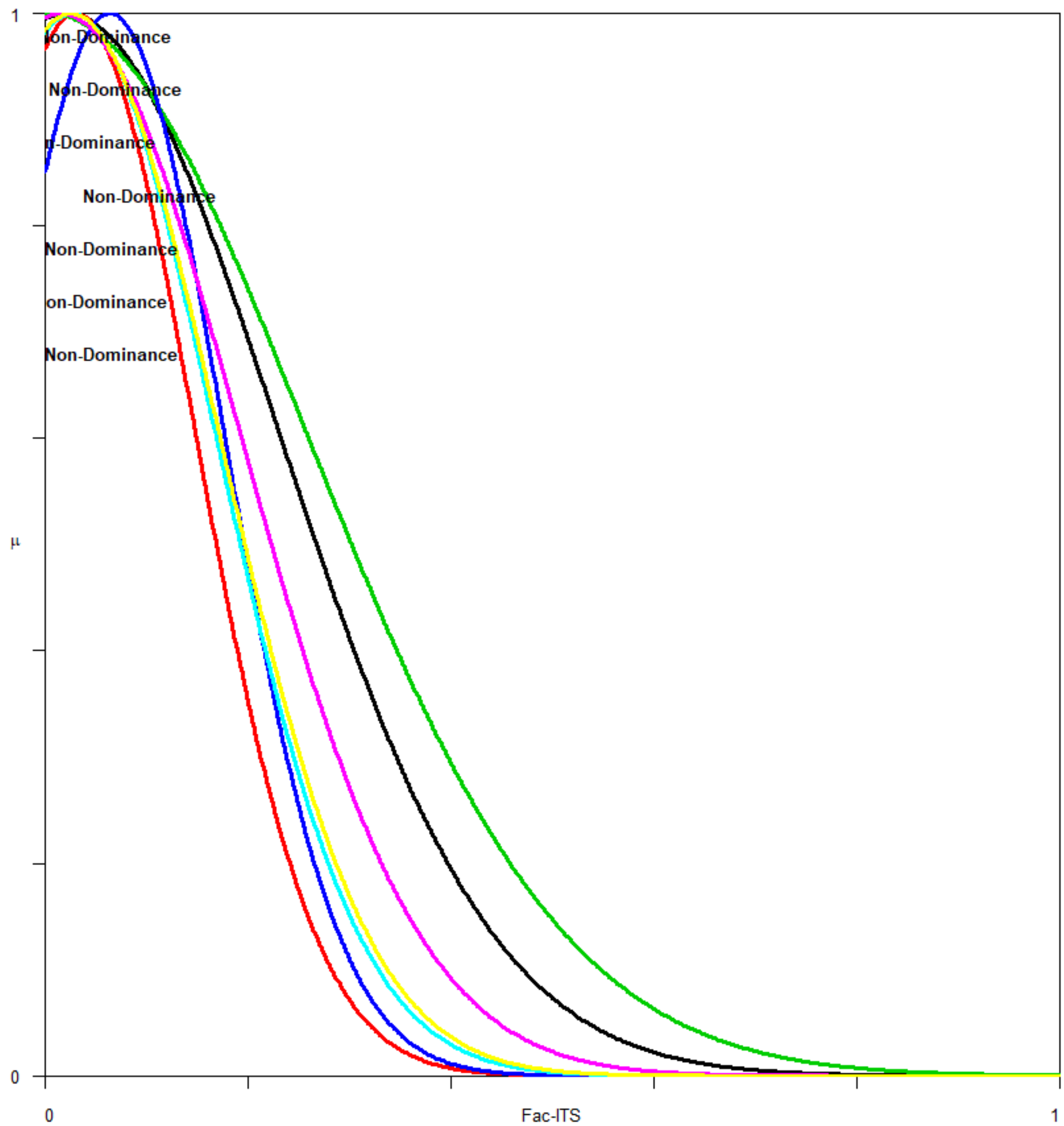

Figure B.4. Membership Function of Fac-ITS

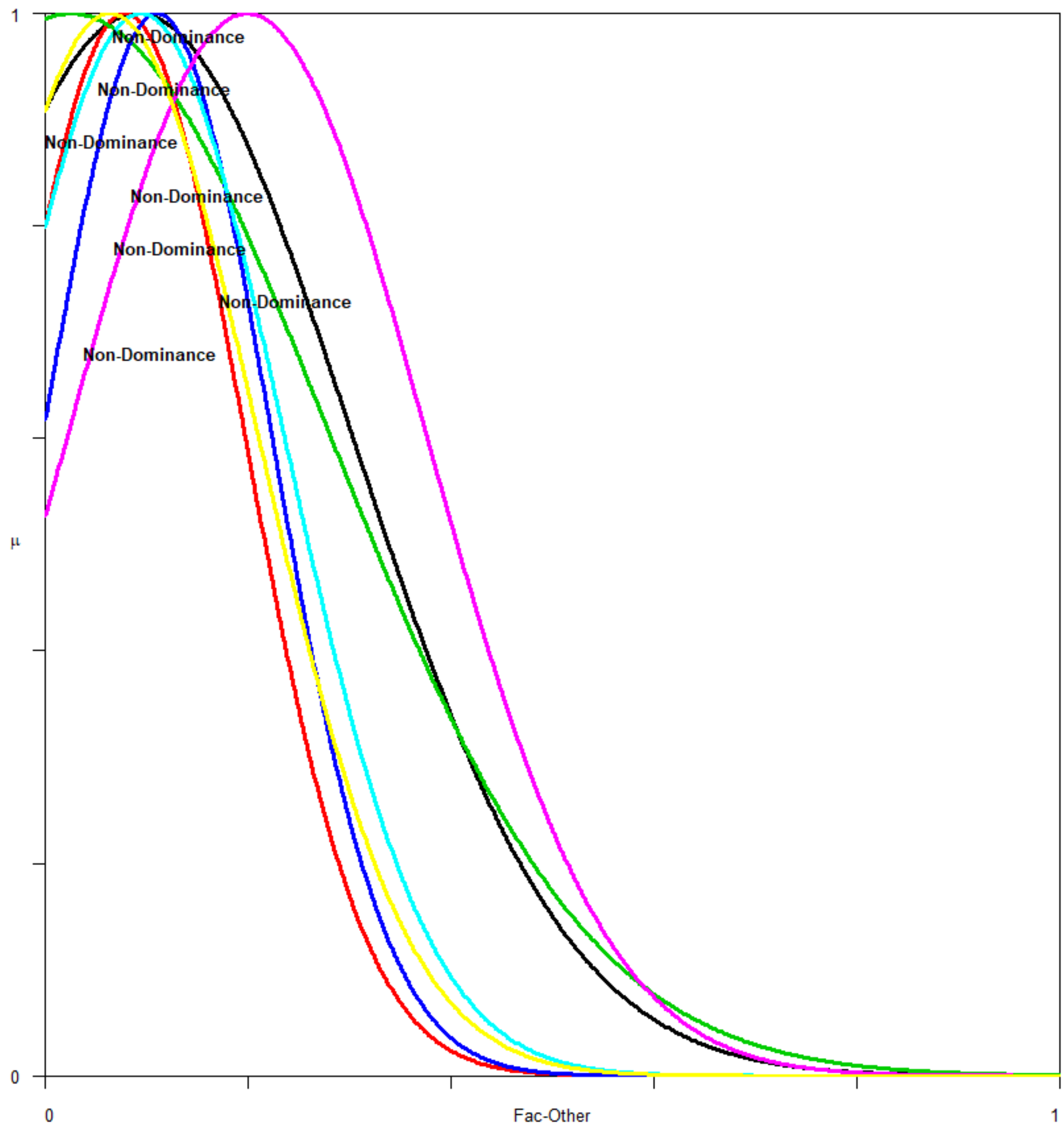

Figure B.5. Membership Function of Fac-Other

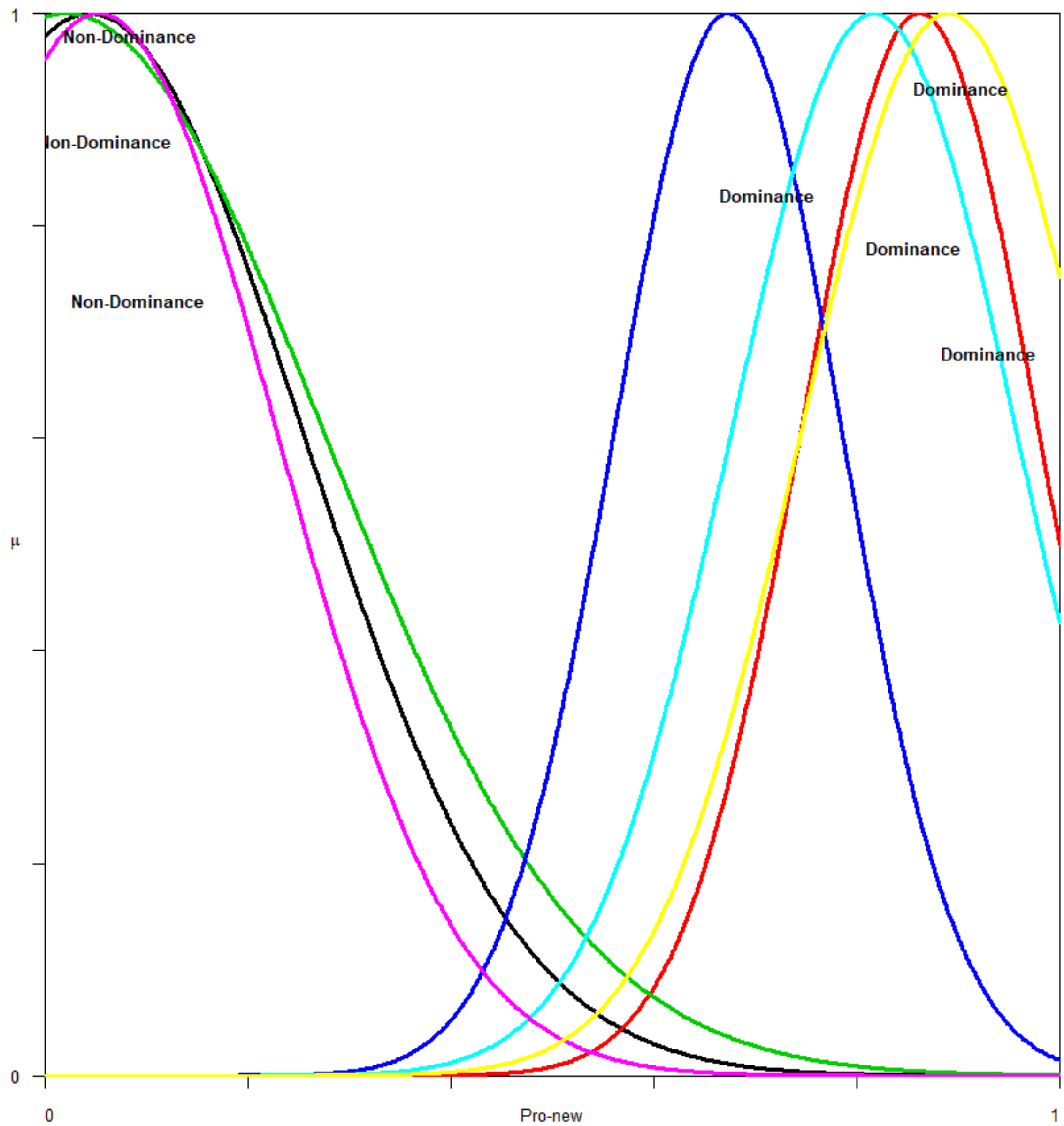

Figure B.6. Membership Function of Pro-New

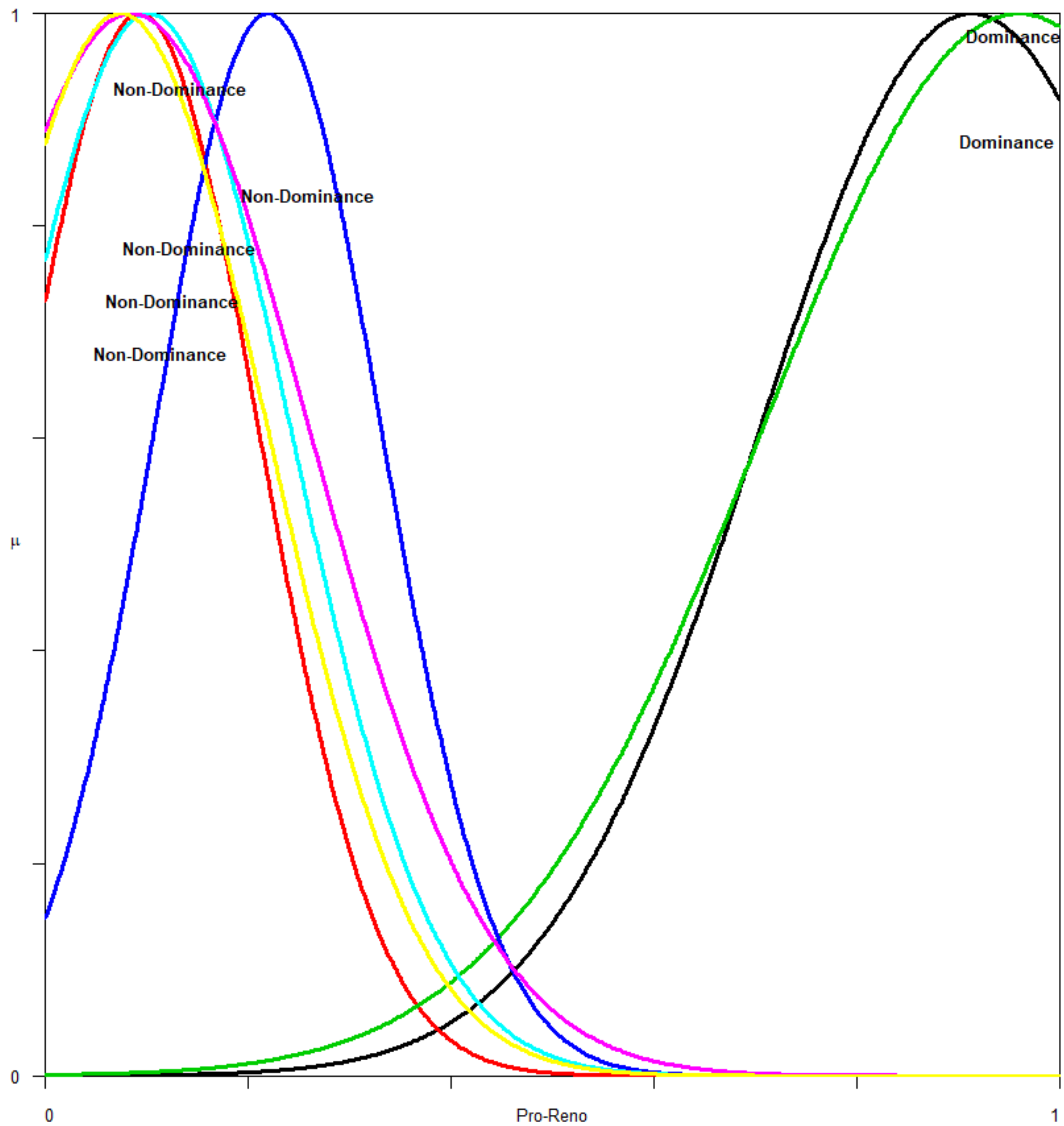

Figure B.7. Membership Function of Pro-Reno

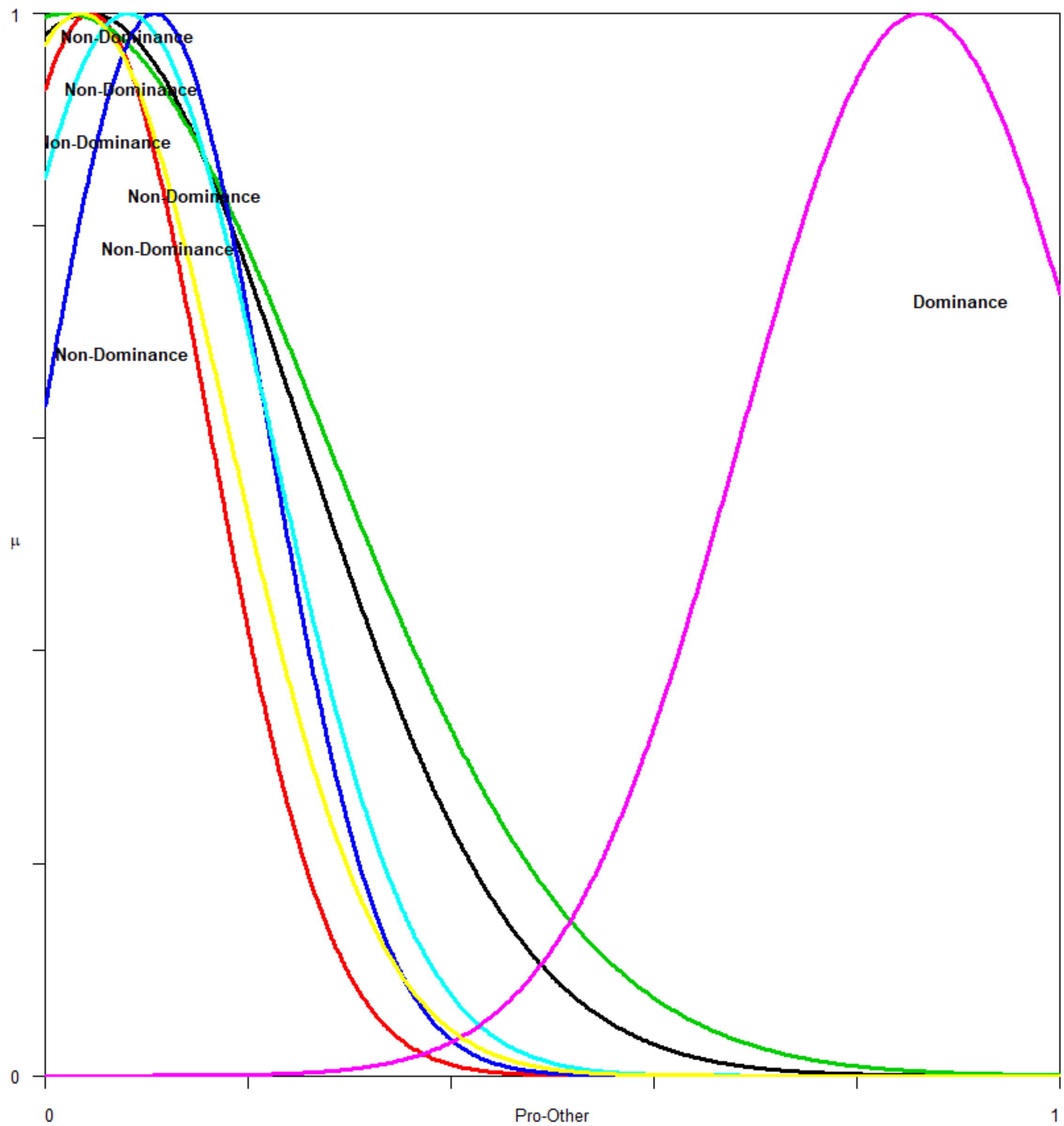

Figure B.8. Membership Function of Pro-Other

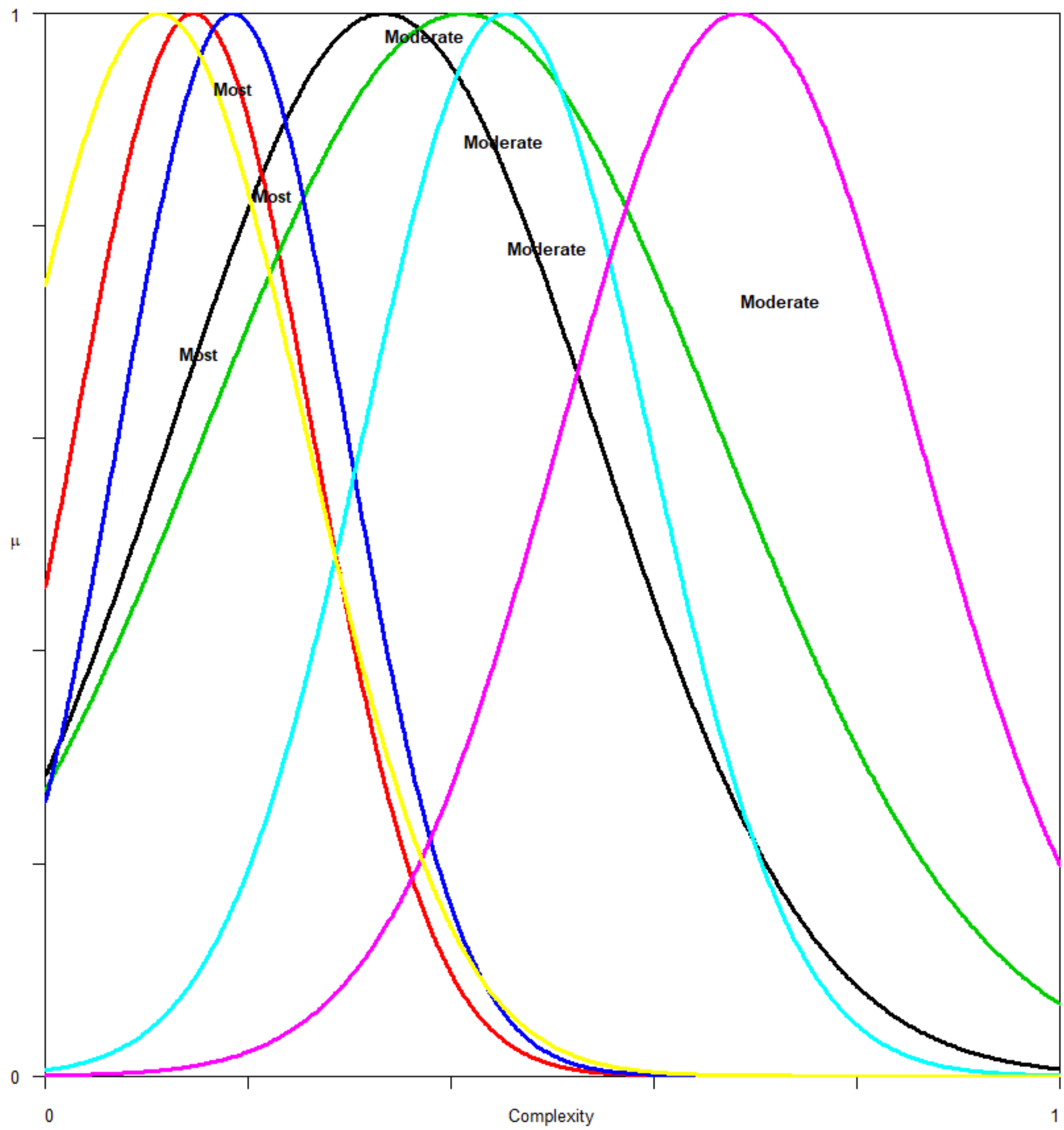

Figure B.9. Membership Function of Project Complexity

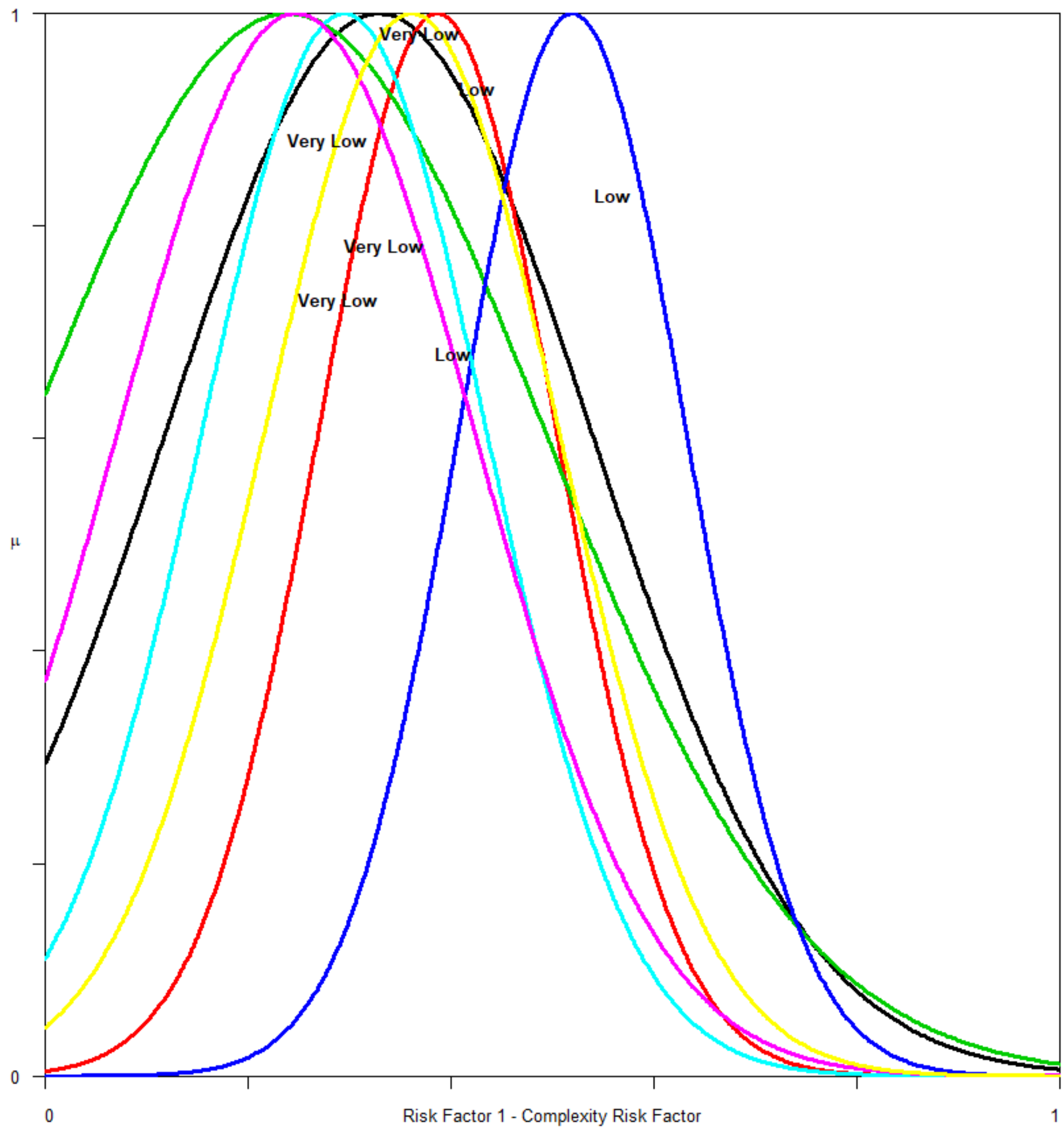

Figure B.10. Membership Function of Risk Factor 1 – Complexity Risk Factor

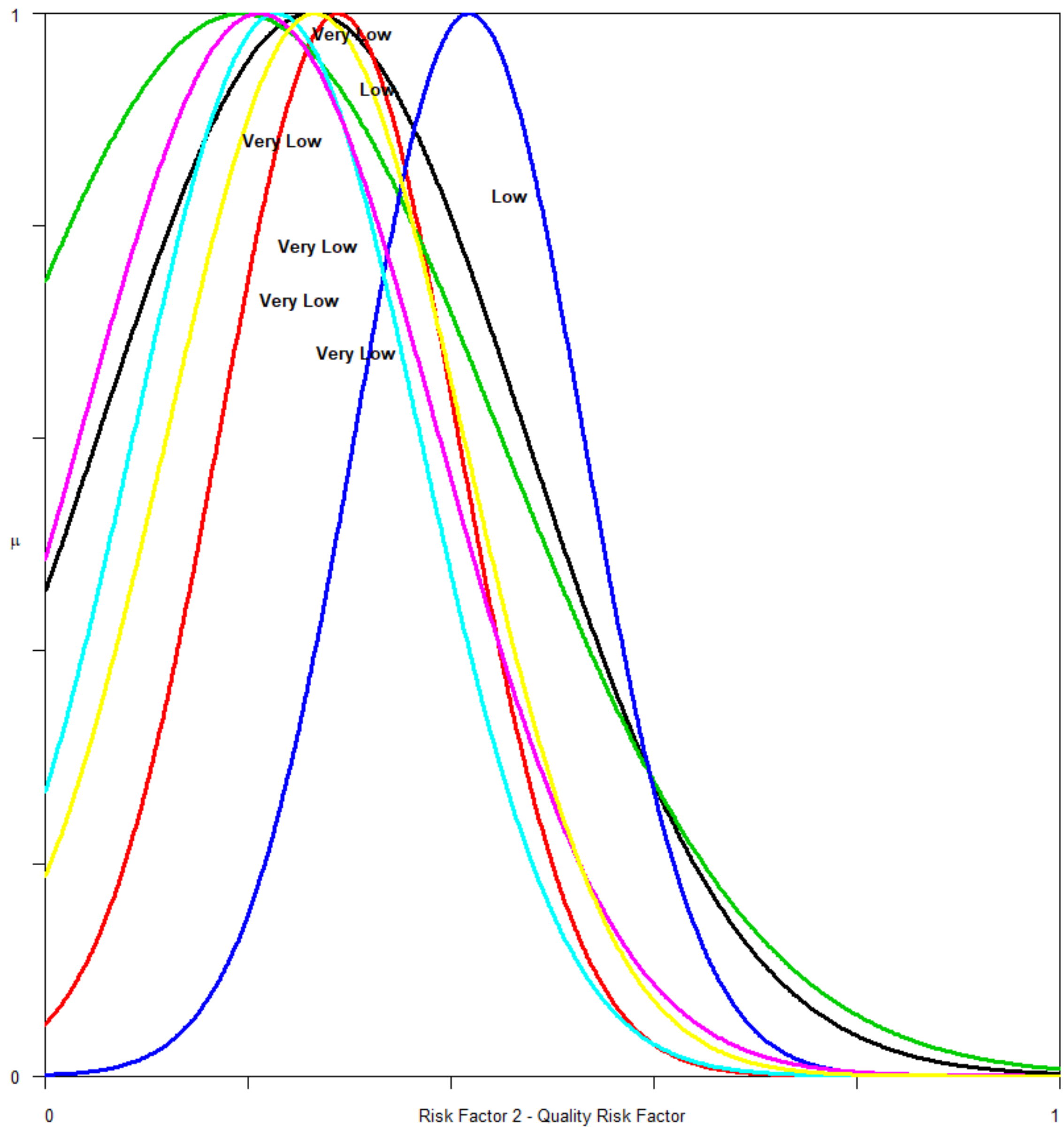

Figure B.11. Membership Function of Risk Factor 2 – Quality Risk Factor

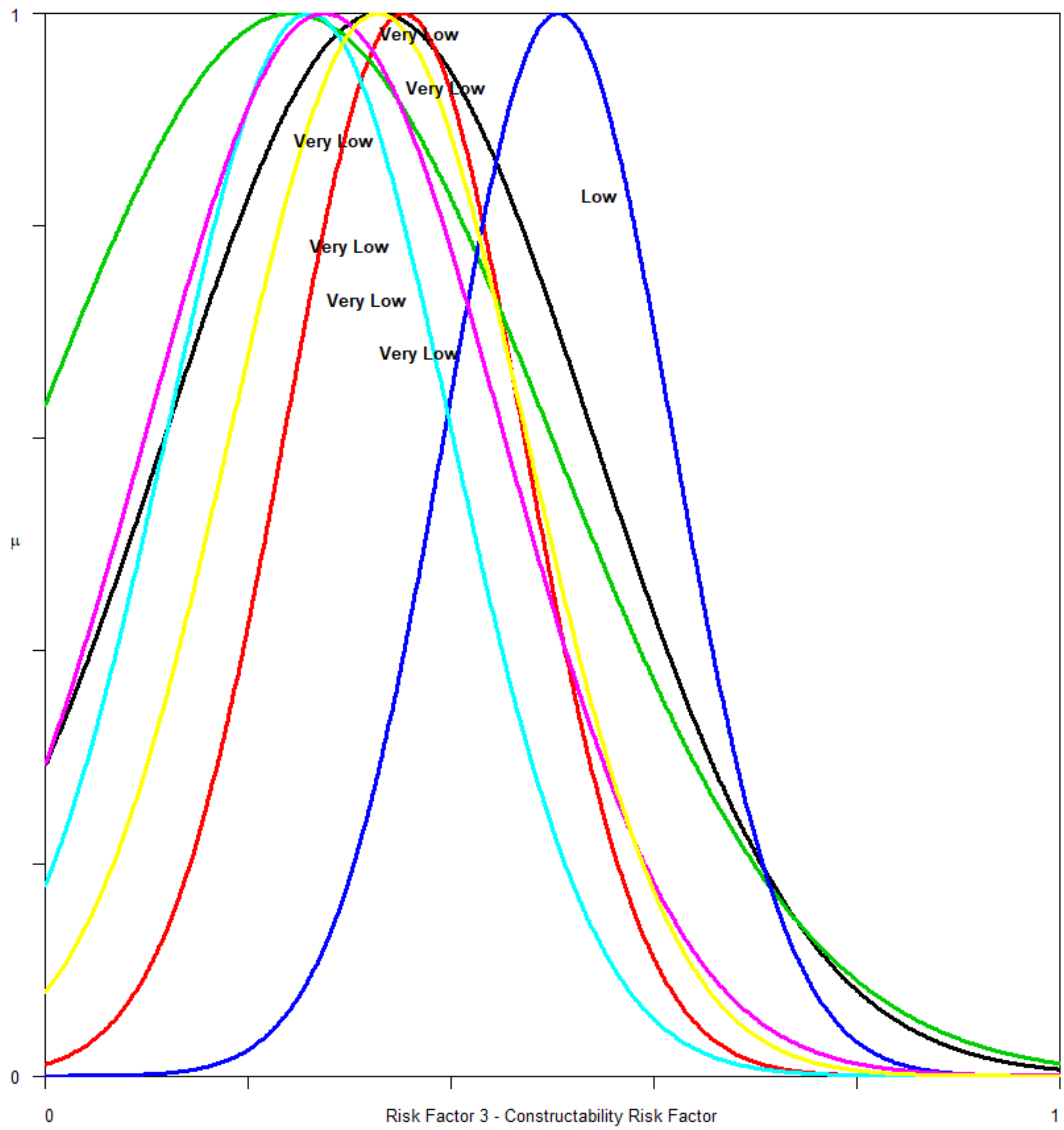

Figure B.12. Membership Function of Risk Factor 3 – Constructability Risk Factor

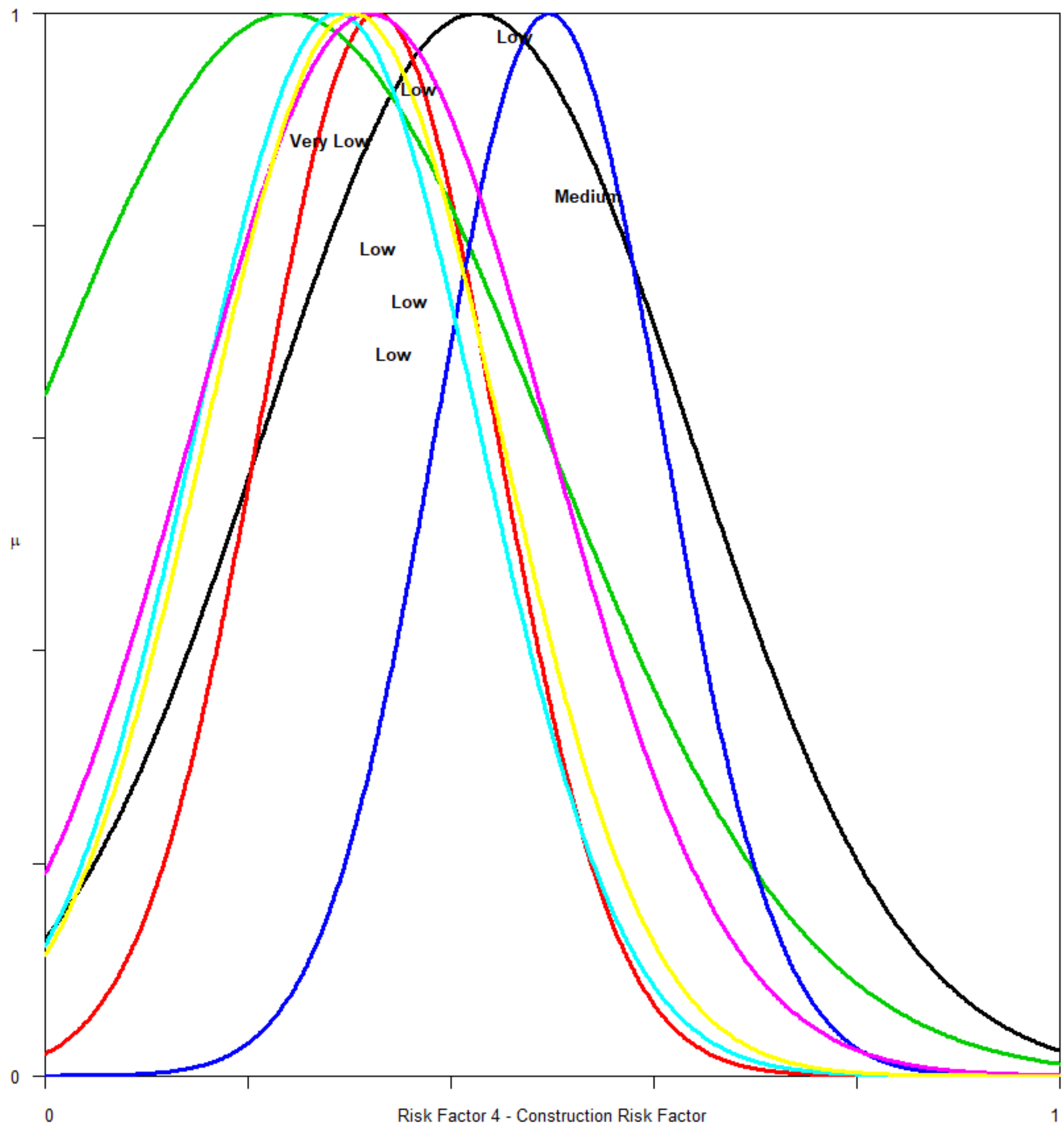

Figure B.13. Membership Function of Risk Factor 4 – Construction Risk Factor

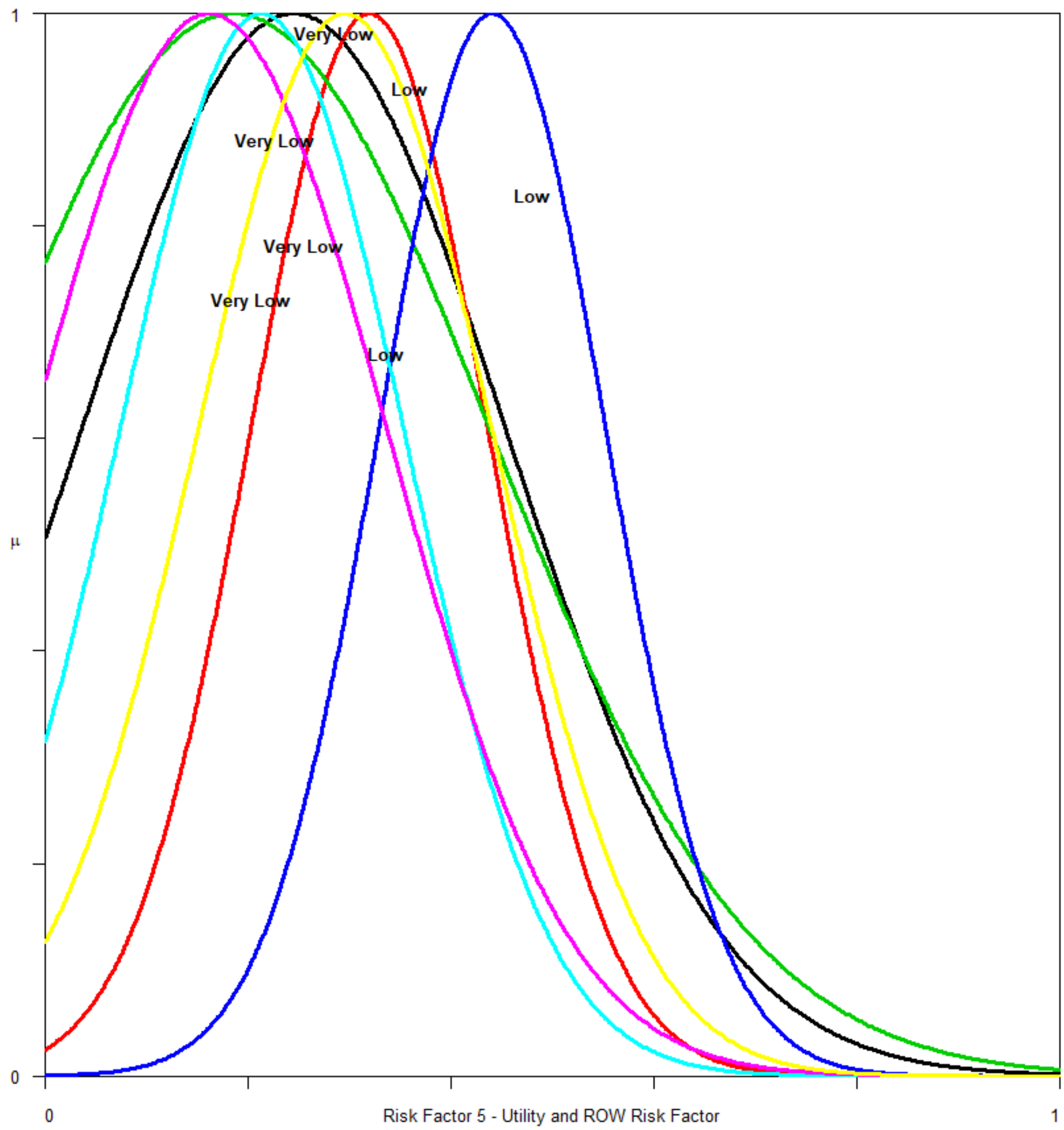

Figure B.14. Membership Function of Risk Factor 5 – Utility and ROW Risk Factor

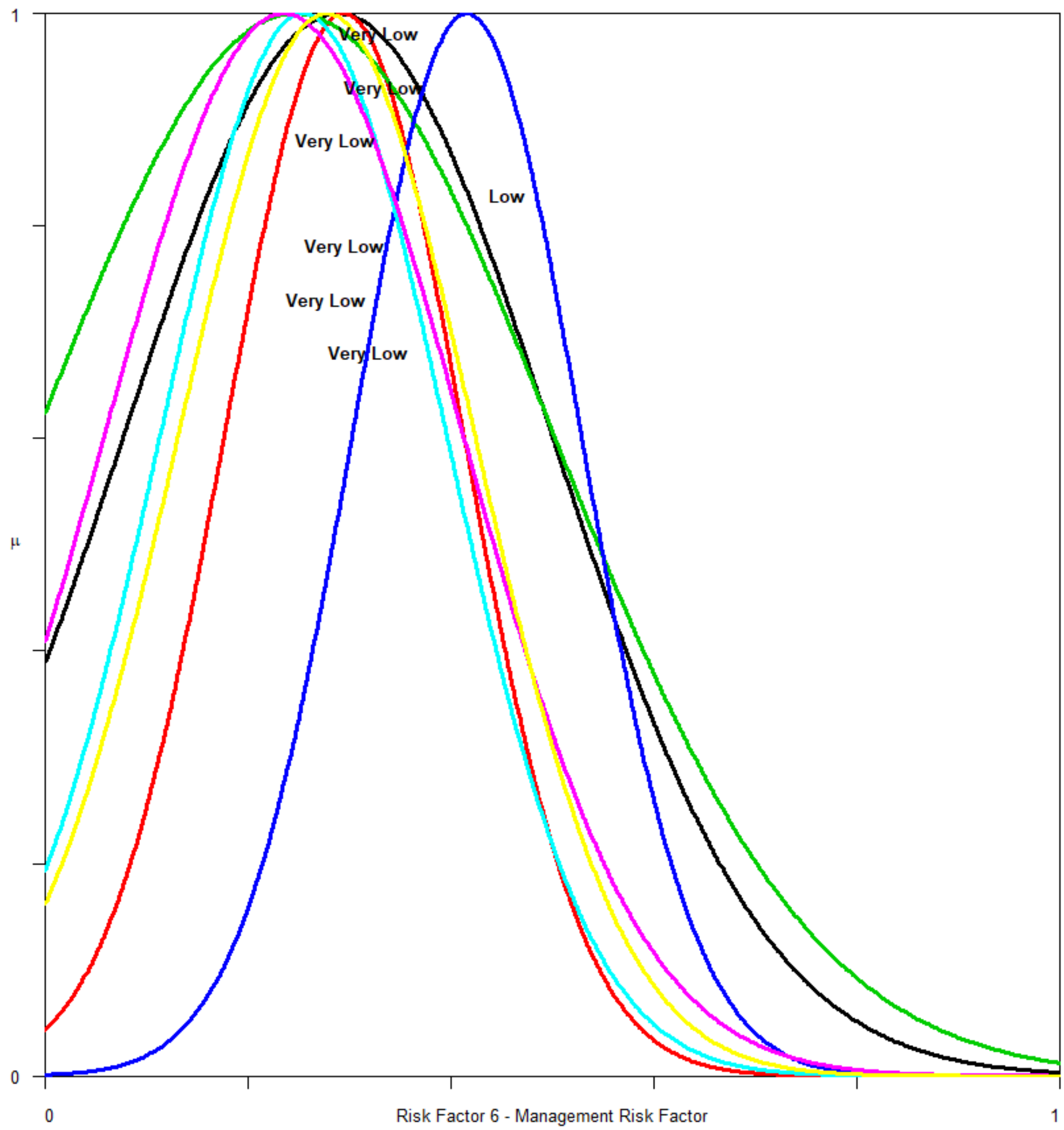

Figure B.15. Membership Function of Risk Factor 6 – Management Risk Factor

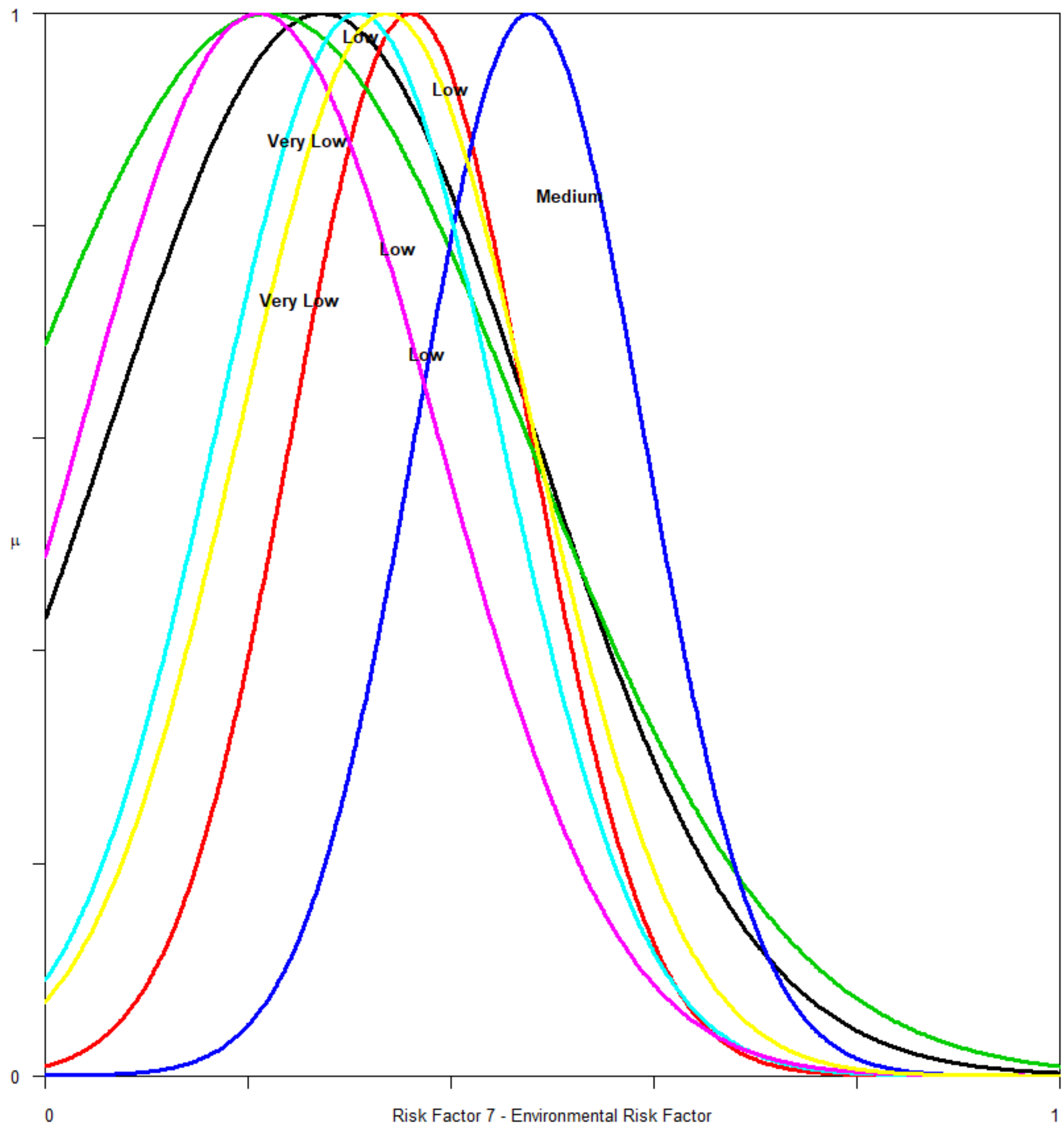

Figure B.16. Membership Function of Risk Factor 7 – Environmental Risk Factor

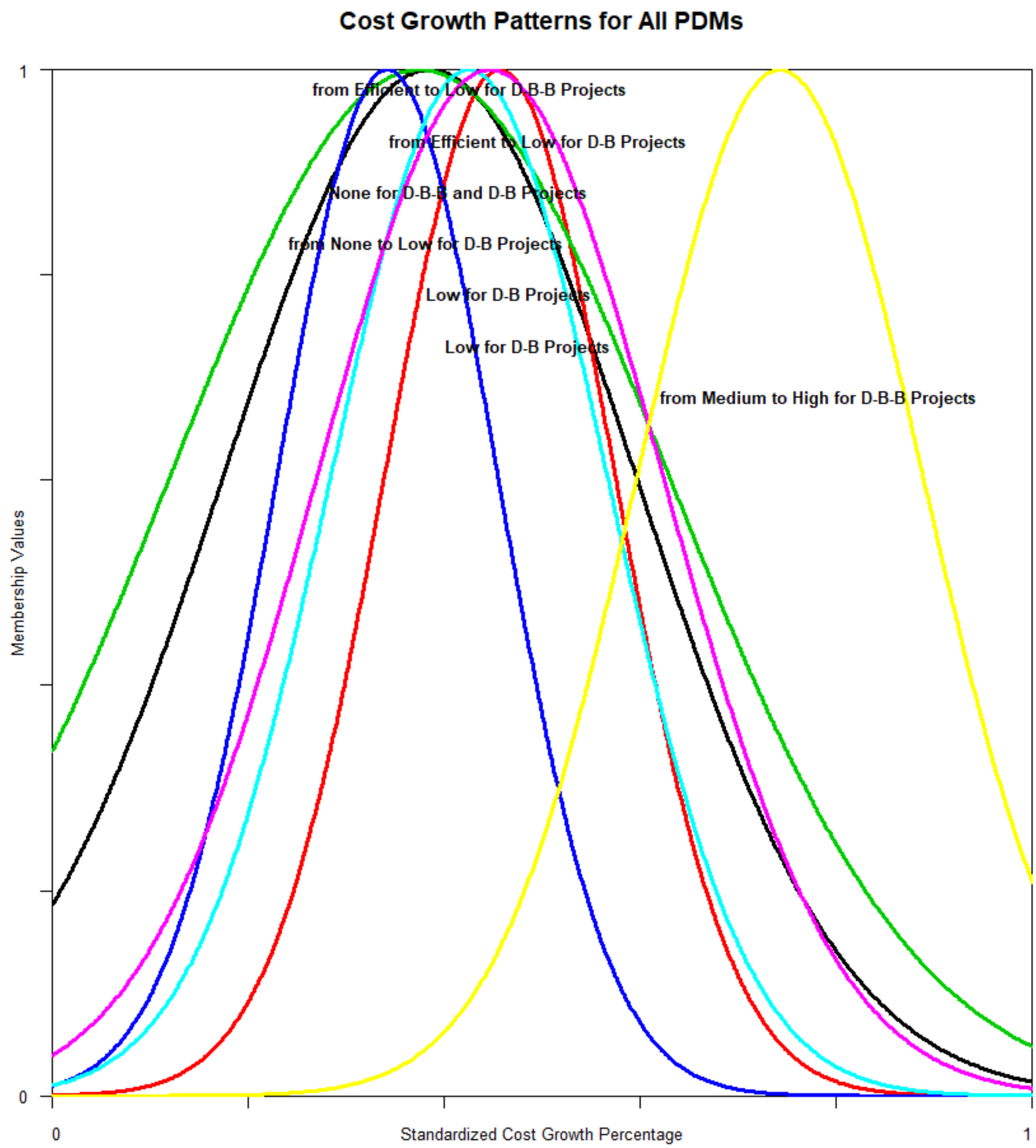

Figure B.17. Membership Function of Cost Growth Patterns

B.2. Rule-Based Inference Formulation

<u>Rule 1</u>: If (Fac-Road is Dominance) and (Fac-Bridge is Non-Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Non-Dominance) and (Pro-Reno is Dominance) and (Pro-Other is Non-Dominance) and (Complexity is Moderate) and (Risk Factor 1 - Complexity Risk Factor is Very Low) and (Risk Factor 2 - Quality Risk Factor is Very Low) and (Risk Factor 3 - Constructability Risk Factor is Very Low) and (Risk Factor 4 - Construction Risk Factor is Low) and (Risk Factor 5 - Utility and ROW Risk Factor is Very Low) and (Risk Factor 6 - Management Risk Factor is Very Low) and (Risk Factor 7 - Environmental Risk Factor is Low) then **(Cost Growth Percentage is from Efficient to Low for D-B-B Projects)**.

<u>Rule 2</u>: If (Fac-Road is Dominance) and (Fac-Bridge is Non-Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Dominance) and (Pro-Reno is Non-Dominance) and (Pro-Other is Non-Dominance) and (Complexity is Most) and (Risk Factor 1 - Complexity Risk Factor is Low) and (Risk Factor 2 - Quality Risk Factor is Low) and (Risk Factor 3 - Constructability Risk Factor is Very Low) and (Risk Factor 4 - Construction Risk Factor is Low) and (Risk Factor 5 - Utility and ROW Risk Factor is Low) and (Risk Factor 6 - Management Risk Factor is Very Low) and (Risk Factor 7 - Environmental Risk Factor is Low) then **(Cost Growth Percentage is from Efficient to Low for D-B Projects)**.

<u>Rule 3</u>: If (Fac-Road is Non-Dominance) and (Fac-Bridge is Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Non-Dominance) and (Pro-Reno is Dominance) and (Pro-Other is Non-Dominance) and (Complexity is Moderate) and (Risk Factor 1 - Complexity Risk Factor is Very Low) and (Risk Factor 2 - Quality Risk Factor is Very Low) and (Risk Factor 3 - Constructability Risk Factor is Very Low) and (Risk Factor 4 - Construction Risk Factor is Very Low) and (Risk Factor 5 - Utility and ROW Risk Factor is Very Low) and (Risk Factor 6 - Management Risk Factor is Very Low) and (Risk Factor 7 - Environmental Risk Factor is Very Low) then **(Cost Growth Percentage is None for D-B-B and D-B Projects)**.

Rule 4: If (Fac-Road is Dominance) and (Fac-Bridge is Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Dominance) and (Pro-Reno is Non-Dominance) and (Pro-Other is Non-Dominance) and (Complexity is Most) and (Risk Factor 1 - Complexity Risk Factor is Low) and (Risk Factor 2 - Quality Risk Factor is Low) and (Risk Factor 3 - Constructability Risk Factor is Low) and (Risk Factor 4 - Construction Risk Factor is Medium) and (Risk Factor 5 - Utility and ROW Risk Factor is Low) and (Risk Factor 6 - Management Risk Factor is Low) and (Risk Factor 7 - Environmental Risk Factor is Medium) then **(Cost Growth Percentage is from None to Low for D-B Projects)**.

Rule 5: If (Fac-Road is Non-Dominance) and (Fac-Bridge is Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Dominance) and (Pro-Reno is Non-Dominance) and (Pro-Other is Non-Dominance) and (Complexity is Moderate) and (Risk Factor 1 - Complexity Risk Factor is Very Low) and (Risk Factor 2 - Quality Risk Factor is Very Low) and (Risk Factor 3 - Constructability Risk Factor is Very Low) and (Risk Factor 4 - Construction Risk Factor is Low) and (Risk Factor 5 - Utility and ROW Risk Factor is Very Low) and (Risk Factor 6 - Management Risk Factor is

Very Low) and (Risk Factor 7 - Environmental Risk Factor is Low) then **(Cost Growth Percentage is Low for D-B Projects)**.

<u>Rule 6</u>: If (Fac-Road is Dominance) and (Fac-Bridge is Non-Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Non-Dominance) and (Pro-Reno is Non-Dominance) and (Pro-Other is Dominance) and (Complexity is Moderate) and (Risk Factor 1 - Complexity Risk Factor is Very Low) and (Risk Factor 2 - Quality Risk Factor is Very Low) and (Risk Factor 3 - Constructability Risk Factor is Very Low) and (Risk Factor 4 - Construction Risk Factor is Low) and (Risk Factor 5 - Utility and ROW Risk Factor is Very Low) and (Risk Factor 6 - Management Risk Factor is Very Low) and (Risk Factor 7 - Environmental Risk Factor is Very Low) then **(Cost Growth Percentage is Low for D-B Projects)**.

<u>Rule 7</u>: If (Fac-Road is Dominance) and (Fac-Bridge is Non-Dominance) and (Fac-Drainage is Non-Dominance) and (Fac-ITS is Non-Dominance) and (Fac-Other is Non-Dominance) and (Pro-new is Dominance) and (Pro-Reno is Non-Dominance) and (Pro-Other is Non-Dominance) and (Complexity is Most) and (Risk Factor 1 - Complexity Risk Factor is Low) and (Risk Factor 2 - Quality Risk Factor is Very Low) and (Risk Factor 3 - Constructability Risk Factor is Very Low) and (Risk Factor 4 - Construction Risk Factor is Low) and (Risk Factor 5 - Utility and ROW Risk Factor is Low) and (Risk Factor 6 - Management Risk Factor is Very Low) and (Risk Factor 7 - Environmental Risk Factor is Low) then **(Cost Growth Percentage is from Medium to High for D-B-B Projects)**.

B.3. R-Codes of Fuzzy Rule-Based Inference System

```
### Insert a membership function ###
fis.cg <- newfis(fisName = "Cost Growth Pattern Recognition for All PDMs")
## Var1 - Fac-Road ##
fis.cg <- addvar(fis = fis.cg, varType = "input", varName = "Fac-Road", varBounds = c(0, 1) , method = "fuzzification") # , params = c(100,0)
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Dominance", mfType = "gaussmf", mfParams = c(0.2097849, 0.5913866)) # Cluster
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Dominance", mfType = "gaussmf", mfParams = c(0.1173441, 0.6298234)) # Cluster
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Non-Dominance", mfType = "gaussmf", mfParams = c(0.2536496, 0.1185072)) # Clus
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Dominance", mfType = "gaussmf", mfParams = c(0.1122331, 0.4145175)) # Cluster
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Non-Dominance", mfType = "gaussmf", mfParams = c(0.1403817, 0.2812729)) # Clus
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Dominance", mfType = "gaussmf", mfParams = c(0.1757688, 0.6551783)) # Cluster
fis.cg <- addmf(fis = fis.cg, varType = "input", varIndex =  1, mfName =  "Dominance", mfType = "gaussmf", mfParams = c(0.145321, 0.6796874)) # Cluster 1
```

Figure B.18. Example of R Codes of Gaussian Membership Function

```
# creates a value for dividing the data into train and test. In this case the value is defined as 90% of the number of rows in the dataset.
smp_siz <- floor(0.9*nrow(new1))
# Randomly identifies therows equal to sample size from  all the rows of Smarket dataset and stores the row number in train_ind
train_ind <- sample(seq_len(nrow(new1)),size = smp_siz)
new1.train <- new1[train_ind,] # creates the training dataset with row numbers stored in train_ind
new1.validation<- othersets[validation_ind,] # creates the validation dataset
new1.test <- othersets[-validation_ind,] # creates the testing dataset
newdata.train <- new1.train[order(Cluster),]
newdata.validation <- new1.validation[order(Cluster),]
newdata.test <- new1.test[order(Cluster),]
## 10-Fold Cross-validation
#Randomly shuffle the data
new1.cross <- new1[sample(nrow(new1)),]

#Create 10 equally size folds
folds <- cut(seq(1,nrow(new1.cross)),breaks=10,labels=FALSE)
#Perform 10 fold cross validation
for(i in 1:10){
  #Segement your data by fold using the which() function
  testIndexes <- which(folds==i,arr.ind=TRUE)
  testData <- new1.cross[testIndexes, ]
  trainData <- new1.cross[-testIndexes, ]
}
```

Figure B.19. Example of R Codes of K-fold Cross-Validation (Training, Validation, and Testing Datasets)

```r
## Inserts a rule ##
ruleList <- rbind(c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1), c(2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1,1), c(3,3,3,3,3,3,3,3,3,3
# ruleList <- rbind(c(1,1,1,1,1,1,1,1,1,1,1,1), c(2,2,2,2,2,2,2,2,2,2,1,1), c(3,3,3,3,3,3,3,3,3,3,1,1), c(4,4,4,4,4,4,4,4,4
fis.cg <- addrule(fis.cg, ruleList)
showrule(fis.cg)
```

Figure B.20. Example of R Codes of Establishing Seven Rules for Fuzzy Inference System

```r
output$down2 <- downloadHandler(
  filename =  function() {
    paste("MFs", "pdf", sep=".")
  },
  # content is a function with argument file. content writes the plot to the device
  content = function(file) {
    if(NumInput==1){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`1`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
    }else if (NumInput==2){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`2`,input$`3`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
    }else if (NumInput==3){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`4`,input$`5`,input$`6`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
    }else if (NumInput==4){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`7`,input$`8`,input$`9`,input$`10`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
    }else if (NumInput==5){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`11`,input$`12`,input$`13`,input$`14`,input$`15`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
    }else if (NumInput==6){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`16`,input$`17`,input$`18`,input$`19`,input$`20`,input$`21`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
    }else if (NumInput==7){
      out_name = fis$output[[1]]$name
      evalfis(c(input$`22`,input$`23`,input$`24`,input$`25`,input$`26`,input$`27`,input$`28`),fis)
      pdf(file)
      plot(D_x,D_y, type="l", main=c(out_name,D_out), col="blue", lwd=2)
      dev.off()
```

Figure B.21. Example of R Codes of Fuzzy Inferencing Process