Developing and Optimizing Context-Specific and Universal Construction Labour Productivity Models

by

Abraham Assefa Tsehayae

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

Doctor of Philosophy

in

Construction Engineering and Management

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Construction labour productivity (CLP) significantly influences the profitability of construction companies; however, CLP exhibits the highest variability among project resources and is a major source of project risk. The construction industry is thus constantly searching for ways to improve labour productivity. Unfortunately, despite long-term, continued research and industry practice, predicting and improving CLP remains a challenge. Previous productivity studies mainly focus on factor and activity models, using factor models to model productivity with context-specific influencing parameters (factors and practices), and activity models to model the relationship between productivity and work sampling proportions (WSP). However, modeling CLP remains a challenge as for a given context, the complex impact of the multiple subjective and objective variables, made up of critical factors, practices, and WSP; have to be considered simultaneously, while maintaining a high accuracy and interpretability in developed models. To address these challenges, this thesis presents advanced frameworks for the development of a series of interpretable and accurate fuzzy inference based context-specific CLP models, which are then abstracted to develop the universal CLP models, and facilitate a better understanding of the variables that influence CLP.

The development of the CLP models included identifying, classifying, quantifying, and documenting the variables influencing CLP. By analyzing existing literature in the field of CLP analysis and modeling, the influencing variables, made up of 169 parameters and 7 work sampling categories, were identified and quantified. The research conducted extensive field data collection from 11 construction projects across Alberta, Canada, spanning over a time period of 29-months; and documented information using factor survey, factors and practices documentation, work sampling studies, foreman delay surveys, craftsman questionnaires, and productivity measurements.

First, the research identified the key variables influencing CLP using expert and data-driven approaches in order to reduce the large feature space of the variables. Next, the role of work sampling proportions in CLP modeling was formulated by testing the fundamental assumption of activity models—that CLP improves if more time is spent on direct work activities—and analysis results showed that using

work sampling proportions alone, it is not possible to accurately predict CLP. Thus, a system-based modeling framework to incorporate work sampling proportions with factors and practices leading to improved CLP modeling and analysis was developed. Then, an operational definition of context for CLP modeling was formulated and associated context attributes were developed, based on the 5W1H (Who, What, Where, When, Why, and How) question and answers approach, and employed together with the system-based CLP modeling framework for the development of a series of context-specific CLP models after combining projects sharing similar contexts. Using a hybrid fuzzy multi-objective optimization framework, the learning ability of the developed fuzzy inference system CLP models was improved. Finally, a context adaptation framework for transferring knowledge among contexts was developed using linear and non-linear adaptation on the membership functions of the context-specific fuzzy CLP models, and a framework for developing universal CLP models is established.

The main contributions of this research to the state of art of CLP modeling and analysis are: (1) evaluation of the usefulness of relying on work sampling proportions like direct work or tool time to predict CLP, (2) development of a system model framework for CLP, which provides a better understanding of CLP and the variables influencing CLP, (3) addressing the challenges faced in past CLP models by developing and optimizing fuzzy inference CLP models, (4) presenting an operational definition of context for CLP modeling for characterizing and classifying construction projects and assisting in the process of grouping similar projects for more accurate context-specific CLP model development, and (5) developing frameworks for adaptation and abstraction of context-specific CLP models. The developed frameworks and findings of this study are of a value to researchers and industry practitioners and provide a better understanding of CLP, the variables influencing CLP, and how work-study methods like work sampling can be integrated to provide an accurate CLP analysis tool.

Preface

This thesis is an original work by Abraham Assefa Tsehayae. The research project, of which this dissertation is based on, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Critical factors for craft work improvement on Alberta construction projects”, Study ID: Pro00023676, approved on June 30, 2011. The following researchers in the Hole School of Construction Engineering at the University of Alberta: Mr. Moataz Omar, Mr. Mohammad Raoufi, Mr. Farhad Shams, Dr. Adel Awad, Mr. Nima Gerami Seresht, Mr. Drew Delbaere, and Mr. Nasir Siraj helped with data collection.

Chapter 3 of this thesis has been published as Tsehayae, A.A., and Fayek, A. Robinson (2014). “Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects.” *Can. J. Civil Eng.*, 41 (10), 878-891. Chapter 4 of this thesis has been submitted for publication as Tsehayae, A.A., and Fayek, A. Robinson (2015). “System model for analysing construction labour productivity.” *J. Constr. Innov.: Info. Proc. Manage.,* in review, submitted July 28, 2015. Chapter 5 of this thesis has been submitted for publication as Tsehayae, A.A., and Fayek, A. Robinson (2015). “Developing and optimizing context specific construction labor productivity models.” *J. Constr. Eng. Manage.*, in review, submitted June 30, 2015. I was responsible for the major parts of the data collection and analysis as well as the composition of the three manuscripts. A. Robinson Fayek was the supervisory author and was involved with concept formation and composition of the three manuscripts.

Dedication

I dedicate this thesis to my mother, **Lemlem Baraky Haile**. Her love, support, and encouragement laid the foundation and discipline needed to complete this research work.

Acknowledgements

First of all, I would like to acknowledge my supervisor, Dr. Aminah Robinson Fayek, for her expertise, close follow up, encouragement, and support. Her hard working culture has been a role model for me and inspired me to pursue excellence. My next acknowledgments go to my doctoral committee, Dr. Ming Lu, Dr. Witlod Pedrycz, and Dr. Marwan El-Rich for the guidance and advice along the course of this research. I am very grateful to Dr. Witold Pedrycz, for his valuable ideas and comments during the initial stages of the research.

I would like to express my sincere appreciation to all of the companies that participated in this study for their cooperation, time, and the valuable information they provided. I also particularly recognize Mr. Moataz Omar, Mr. Mohammad Raoufi, Mr. Farhad Shams, Dr. Adel Awad, Mr. Nima Gerami Seresht, Mr. Drew Delbaere, and Mr. Nasir Siraj, researchers in the Hole School of Construction Engineering at the University of Alberta, who diligently helped with data collection. This research was conducted under the NSERC Industrial Research Chair in Strategic Construction Modeling and Delivery, held by Dr. Aminah Robinson Fayek. The financial support of the industrial partners to this Chair and of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

I would like to express my deepest gratitude and sincere thanks to my family, Selam Redda, Henock Assefa, and Romha Assefa for their constant support and encouragements to sustain the pursuit of my PhD study.

Finally, I would like to acknowledge my fellows Nasir Siraj, Daniel Meles, Samson Tassew, Esayas G/Yohannes, Miao Zhen, Naimeh Sadeghi, and Ronald Ekyalimpa for all their help during my studies. I would also like to take this opportunity to thank all the individuals, colleagues, and friends who assisted me during this research project.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABM	Agent based modeling
AI	Artificial intelligence
CFS	Correlation-based feature selection algorithm
CII	Construction Industry Institute
CLP	Construction labour productivity, defined as the ratio of outputs to inputs
CSAM	Context-specific augmented construction labour productivity model
DB	A component of fuzzy inference system that associates membership functions to the linguistic variables used in the rule-base
DEA	Data envelopment analysis
DS	Desired sample size
FCM	Fuzzy C-Means
FIS	Fuzzy inference system based on if-then rules and linguistic variables
GA	Genetic algorithm
I	Input parameters made up of factors and practices influencing CLP
MF	Membership function
NN	Neural network
O	Output parameter representing CLP
P	Process parameters representing work sampling proportions
PM	Project management
PMI	Project Management Institute
RB	A component of fuzzy inference system composed of linguistic if-then rules
RMSE	Root mean square error
WS	Work sampling
WSP	Work sampling proportions
UPM	Universal productivity model

Chapter 1: Introduction[1]

1.1: BACKGROUND

The construction industry is a discipline that attempts to successfully deliver and manage capital facility and infrastructure projects under uncertain, dynamic, and risk-filled environment (PMI 2007). The industry is a vital part of many national economies and in Canada for the last five years, construction, on average, contributed to 7.2% of the Gross Domestic product and provided employment for about 7.6% of the workforce (Statistics Canada 2015a; Statistics Canada 2015b). According to Walker (2015), the industry can be effectively framed as an open system conversion process, where a set of inputs (land, knowledge, information, energy, materials, etc.) are transformed, using labour and/or machines, to outputs (buildings, roads, industrial plants, bridges, etc.). The efficiency of construction systems is measured using construction productivity; consequently, construction productivity has a wide range of applications, each having different meanings and definitions.

Traditionally, research studies have defined productivity to suit a specific purpose for the construction industry, at either, the industry, project, or activity level (Thomas et al. 1990). Productivity can be generally defined as the "amount of goods and services produced by a productive factor in a unit of time" (Drewin 1982). The most common construction productivity metrics are: unit rate (ratio of labour cost to units of output); labour productivity (ratio of work hours to units of output); and productivity factor (ratio of scheduled or planned to actual work hours) (Gouett et al. 2011). The efficiency of activity level systems, focusing on the labour resource of the construction process, is measured using construction labour productivity (CLP). In this thesis, the focus is on CLP, which is defined as the ratio of units of output to units of input work hours—as shown in Eq. (1.1), where higher values are better than lower values.

[1] Parts of this chapter have been published in the Proceedings, ASCE Construction Research Congress 2012, West Lafayette, Indiana, US, May 21-23, pp. 1420-1429; published in Canadian Journal of Civil Engineering, Volume 41, Issue 10, pp. 878-891; and submitted for publication in Journal of Construction Innovation: Information, Process, Management, JCI, 36 manuscript pages, submitted July 28, 2015.

$$\text{Construction labour productivity (CLP)} = \frac{\text{Output (installed quantity)}}{\text{Total labour work-hours}} \tag{1.1}$$

Construction labour productivity, the most commonly used single factor productivity measure, significantly influences the profitability of construction companies, as it represents a significant component of the project cost, ranging from 30-50% (Hanna 2010). However, construction productivity at industry level has stagnated in the Canadian and U.S. construction industries (Harrison 2007; Nasir et al. 2014). CLP also remains low, and is a major source of project risk and exhibits the highest variability among project resources (Moselhi and Khan 2012). The construction industry is thus constantly searching for ways to improve labour productivity. However, before they can propose and implement improvement strategies, industry representatives need an activity-level construction labour productivity model that enables them to fully understand which parameters (factors and practices) cause productivity to change and by how much (Thomas et al. 1990). Such models also play a key role in construction estimating, scheduling, and planning decisions (Yi and Chan 2014). Construction labour productivity, referred to as the output variable (*O*), deals with the efficiency of labour crews in the complex process of converting inputs (labour, material, equipment, etc.) to outputs (project products) in various construction project contexts. CLP is situated in an environment that is more complex and unpredictable than the conversion process itself, causing a number of parameters to either directly or indirectly influence CLP. Different parameters, made up of various factors and practices (e.g., crew size, crew composition, co-operation among craftsperson, location of work scope, complexity of task, weather condition, risk management practice, etc.) are known to affect the conversion process. Of these parameters, this study considers those that critically influence CLP as input variables (*I*) in order to further examine their effects on CLP (Fig. 1.1).

Additionally, understanding how time is used during the input-to-output conversion process is also vital to modeling CLP; work-study methods are commonly employed for this purpose. Work sampling, a method used to determine the amount of time workers spend performing direct (productive) work, handling material, waiting, etc. is the most widely used work-study method (Josephson and Björkman

2013). Work sampling proportions summarize the actual utilization of labour work hours, and in this research are represented as process variables (*P*); they provide an in-depth examination of what happens during the conversion process (Fig. 1.1). In order to improve CLP, appropriate analysis and modeling is required so as to clearly illustrate how input variables affect the efficiency of the conversion process. Such an analysis must establish the relationships between the three *system model variables*—*Input*, *Process*, and *Output* (Fig. 1.1)—so as to examine the cause and effect of the input and process variables on CLP.

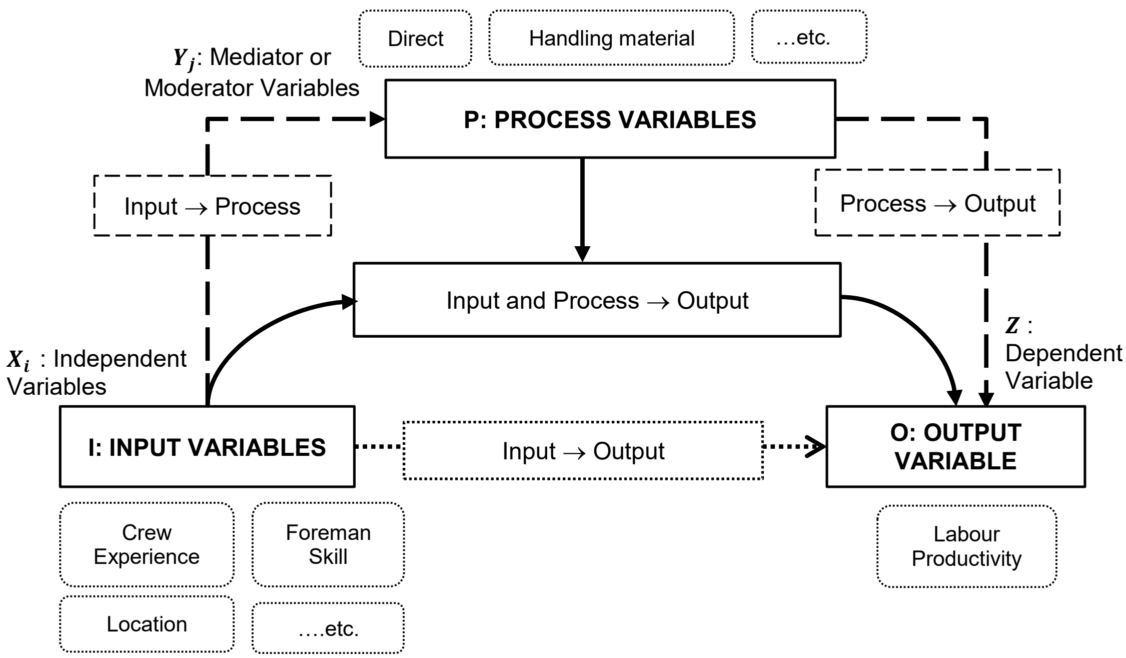

Figure 1.1: Representation of the System Model of CLP as an Open Conversion Process

1.2: Problem Statement

Because of its significance to project performance, CLP has been well studied. Accordingly, several CLP studies have identified numerous parameters that influence CLP (Thomas et al. 1990; Oduba 2002; Liberda et al. 2003; Song and AbouRizk 2008; Dai et al. 2009; Oral et al. 2012; Tsehayae and Fayek 2014; Gerek et al. 2014). However, despite the extensive research in the area, consensus on the classification of parameters and generalization of key parameters is yet to be achieved (Panas and Pantouvakis 2010). Therefore, the *first problem* in CLP modeling is related to the identification of the multilevel, complex, and context-dependent key parameters (factors and practices), influencing CLP in different project contexts. Additionally, numerous CLP models have been developed for analyzing the impact of the influencing parameters on CLP, and have used a variety of modeling techniques (Yi and Chan 2014). Overall, these tested approaches can be categorized as either factor or activity models. Factor models relate the different input variables—made up of key influencing parameters (factors and practices) like crew size, weather condition, etc.—to labour productivity. Activity models mainly relate the process variables, in terms of work sampling proportions, to labour productivity. However, no previous studies have succeeded in developing an integrated system approach by investigating the overall relationship between both input and process variables and CLP. Therefore, the *second problem* in CLP modeling is related to the lack of a system-based CLP modeling framework, which would have enabled the examination of the relationship among the three system variables.

The *third problem* is related to the development of accurate and interpretable CLP models, which remains a challenge due the complex variability of CLP, the limited data availability to study CLP under various contexts, and the requirement of considering the complex impact of the multiple variables simultaneously, while maintaining a high accuracy and interpretability in the developed models. CLP studies have thus focused on the use of artificial intelligence techniques like neural networks and fuzzy inference systems to model CLP (Oral et al. 2012; Fayek and Oduba 2005). Fuzzy inference systems (FISs) are based on fuzzy set theory and if-then rules and have provided effective tools to solving engineering problems in biomedical engineering, robotics, pattern recognition, image processing, and control application areas (Botta 2008), and have the ability to address the identified challenges in modeling CLP. The use of FISs have also been gaining widespread attention in construction research (Chan et al. 2009), however, FISs have had limited application in CLP modeling, and the few studies using FIS had limitations in the development of membership functions and if-then rules from data (Mao 1999; Fayek and Oduba 2005). FISs have the capability to deal with the large number of subjective variables, by means of fuzzy sets representing linguistic terms; model the complexity of CLP using if-then rules, which can be developed using limited data; and are highly interpretable. However, fuzzy inference systems have one significant limitation in that they lack the ability to learn from data and optimize their model parameters, resulting in the *fourth problem* faced in modeling CLP using FIS. FIS-based models contain several model parameters that can be optimized; thus hybridizing FISs through combination with other artificial intelligence techniques has been tested, resulting in improved learning capabilities (Awad and Fayek 2013).

In past CLP studies, the identified influencing parameters and the associated CLP models were context dependent, as the identified parameters and their degree of impact on CLP varied from project to project (Gerek et al. 2014), implying that CLP is a context-sensitive problem, and as such the developed CLP models are specific to the context of development (Thomas et al. 1990). However, only a few CLP studies had an explicit definition of the context of the CLP modeling processes. Most models overlooked the role of context in CLP modeling and its importance for formulating the circumstances that form the setting of the CLP model and its development process (Yi and Chan 2014). The *fifth problem* in CLP modeling is thus the lack of a clear and explicit representation of context in past CLP studies, and evaluation of the usefulness of context to characterize and classify construction projects and assist in the process of grouping similar projects for more accurate CLP model development.

Context plays an essential role in CLP research, as it defines in which scenarios the findings of the CLP models are applicable. Thus, past CLP models can be conceptualized as the observation of activity level construction systems from different points of view or contexts, denoted as Context 1, Context 2,…, Context-p, which are formulated based on the data collected for each context, denoted as Data 1, Data 2,…, Data-p, as shown in Fig. 1.2. However, in CLP modeling field an approach for transferring the knowledge represented in the context-specific CLP models from one context to another is missing. Such an approach is particularly important when modeling new contexts for which data availability is limited; and existing models cannot be applied without some adaptation. The *sixth problem* is thus the unavailability of a context adaptation framework to modify CLP model parameters and enable the transfer of knowledge from one context to another. Additionally, in CLP modeling, a universal model that represents a versatile knowledge that can be used in any context is needed, as developing adequate number of context-specific models representing each unique construction context is difficult to achieve. A universal CLP model represents the generalized context-free knowledge base and can be used to develop best practices and principles for improvement of CLP.Therefore the *seventh problem* in CLP modeling research is the absence of a universal CLP model and an approach for its development.

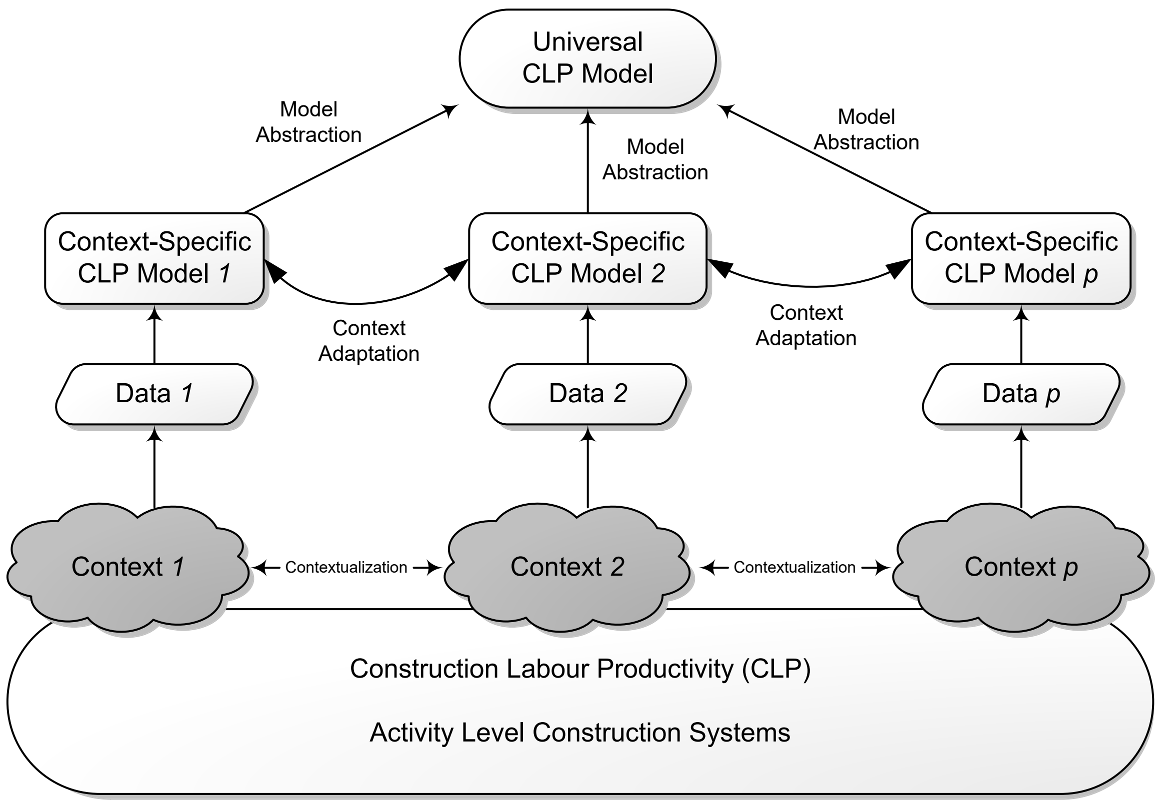

Figure 1.2: The Emergence of Context-Specific and Universal CLP Modeling Approach

1.3: RESEARCH OBJECTIVES

The overall aim of this thesis is to present a methodology for the development of interpretable and accurate context-specific and universal CLP models that facilitate a better understanding of the variables that influence CLP. The methodology examines the effect of the numerous context-sensitive influencing variables, made up of subjective and objective factors, practices, and work sampling proportions causing the complex variability of CLP, using data-driven and optimized fuzzy inference system CLP models. The detailed objectives of the research, addressing the identified problems in CLP modeling, are grouped under the following three main categories:

1. To advance the body of knowledge related to the input and process variables influencing CLP; thereby addressing the *first problem* in CLP modeling using objectives 1 (a)—1(d) and the *second problem* using objectives 1(e) and 1(f):
    a. To identify, classify, and develop a comprehensive hierarchal list of the multilevel and context-dependent parameters (factors and practices) influencing CLP in different project contexts.
    b. To develop a methodology for quantification of subjective and objective parameters influencing CLP.
    c. To identify the most critical context-specific parameters (factors and practices) influencing CLP using two staged approach: (1) expert-driven approach employing context-centered surveys intended to verify the hierarchal list of parameters influencing CLP, establish the existence or frequency of the parameters in studied construction projects, and establish the context-specific nature of key parameters based on positive and negative effects on CLP as reported by respondents from project management and trade groups; and (2) data-driven approach employing a feature selection algorithm on the hierarchal list of parameters and field data collected for respective contexts.
    d. To test the fundamental assumption of activity models—that CLP improves if more time is spent on direct work activities, and evaluate the role of process variables or work sampling proportions in CLP modeling.
    e. To develop a novel system model approach for improved prediction of CLP using input variables made up of key influencing parameters in conjunction with process variables made up of work sampling proportions.

2. To advance the state of the art in the development of accurate and interpretable fuzzy inference system-based CLP models; thereby addressing the *third problem* in CLP modeling using objective 2(a) and the *fourth problem* using objective 2(b):
    a. To develop a methodology for the development of fuzzy inference systems and membership functions using limited data.
    b. To develop a multi-objective optimization framework for improving not only the accuracy, but also the interpretability of fuzzy inference CLP models, and validation of such models using appropriate validation strategies.
3. To develop a novel context-specific and universal CLP modeling methodology providing an improved approach to modeling CLP; thereby addressing the *fifth problem* in CLP modeling using objectives 3(a)—3(c), the *sixth problem* using objective 3(d), and the *seventh problem* using objective 3(e):
    a. To develop an operational definition of context and associated context attributes to explicitly define the context of projects under investigation.
    b. To examine the application of context in CLP modeling by applying the formalized context definition to identify the uniqueness of the studied construction projects.
    c. To investigate the effect of context in CLP modeling using context-specific models, addressing the unique contexts and compare and contrast the performance of context-specific models against a generic CLP model, developed by combining the context-specific data sets.
    d. To develop a context adaptation framework for adapting fuzzy inference CLP models so as to properly adapt the models from one context to another.
    e. To develop an advanced framework for the development of universal CLP model by abstracting context-specific fuzzy models.

1.4: EXPECTED CONTRIBUTIONS

This research described in this thesis is expected to produce the following academic contributions, relevant to academic researchers, and industrial contributions, relevant to industry practitioners.

1.4.1: Academic Contributions

The expected academic contributions of this research are as follows:

- Presenting a comprehensive and hierarchical set of parameters comprised of not only factors but also project practices influencing CLP, thus providing a broader view of CLP and the key issues affecting it.
- Providing a new methodology for quantifying subjective and objective parameters influencing CLP, this is essential for gathering accurate data on parameters influencing CLP.
- Presenting a new hybrid expert and data-driven methodology for evaluating and ranking of the input parameters based on the positive and the negative influence of each distinct parameter on CLP, thereby enabling the identification of enablers as well as barriers to betterment of CLP under different contexts.
- Evaluating the usefulness of relying on work sampling proportions like direct work or tool time to predict CLP, and test the assumption witnessed in CLP research that direct work proportions are highly correlated to CLP.
- Developing a system model framework for CLP, which provides a better understanding of CLP, the parameters influencing CLP, and how work-study methods like work sampling can be integrated to provide an accurate CLP analysis tool.
- Addressing the challenges faced in past CLP models by developing interpretable and accurate fuzzy inference CLP models that explain the impact of multiple subjective and objective variables on CLP, while requiring limited data for development.
- Advancing the state of art in hybrid fuzzy modeling using a genetic algorithm-based optimization process to improve the interpretability and accuracy of developed CLP models.

- Presenting an operational definition of context for CLP modeling and for characterizing and classifying construction projects and assisting the process of grouping similar projects for more accurate context-specific CLP model development.
- Providing a novel context adaptation framework for adapting CLP models from one context to another, thereby facilitating the transfer of knowledge among existing CLP models.
- Providing a novel framework for developing universal CLP models through abstraction of the knowledge bases represented in the context-specific CLP models.

1.4.2: Industrial Contributions

The expected industrial contributions of this research are as follows:

i. Establishing a multilevel factors and practices list affecting construction labour productivity of construction projects.
ii. Identifying and comparing key parameters influencing CLP in building and industrial project contexts in Alberta, Canada, which will provide useful insight on the issues to focus on during construction planning and execution phases.
iii. Offering measurement scales for documenting subjective and objective parameters influencing CLP, and presenting a comprehensive data collection protocol, which provides detailed guidelines for carrying out labour productivity improvement studies.
iv. Developing a Productivity Database to facilitate data collection and analysis, useful for industry practitioners carrying out productivity improvement studies.
v. Developing a tool for predicting CLP for use in construction project cost estimation and scheduling, and developing CLP improvement strategies by identifying optimum values of influencing variables leading to better CLP values.

1.5: RESEARCH METHODOLOGY

The research work presented in this thesis is conducted in four main stages, which are described in the following subsections:

1.5.1: The First Stage

The development of the context-specific and universal CLP models begins with identifying, classifying, quantifying, and documenting the input and process parameters influencing CLP. By analyzing existing literature in the field of CLP analysis and modeling, the input and process parameters are identified and classified into a hierarchal list. Next, quantification of the numerous subjective and objective parameters influencing CLP is carried out. A detailed data collection protocol is also developed to facilitate data collection by several different collectors and to ensure the validity of the data collected from a number of projects. Additionally, a custom-made, server-based database, called ProductivityTracker©, is established to store the vast amount of gathered data and facilitate further modeling steps.

1.5.2: The Second Stage

The large input parameters feature space, made up of the influencing factors and practices, had to be reduced to maintain the interpretability and accuracy of the CLP models. The reduction of the feature space is carried out by identifying the key input parameters (factors and practices) influencing CLP using a hybrid expert and data-driven approaches. A methodology based on factor surveys for collecting expert opinions from different contexts is developed using two survey forms, namely the project management survey and the trade survey. Based on survey responses from project management and trade level project participants, categorized under building and industrial contexts, the key parameters influencing CLP positively and negatively are identified. The internal consistency of the survey responses were examined using Cronbach's alpha values and using statistical analysis the difference in perspective between contexts and also respondents were examined. Using the expert-driven approach, the developed hierarchal list of parameters influencing CLP is verified and the context-specific nature of key parameters is established. A data-driven methodology based on feature selection technique is also employed and the most critical parameters are identified using field data collected for each of the hierarchal parameters. Feature selection was carried out using the Waikato Environment of Knowledge Analysis (WEKA) tool and correlation-based feature selection (CFS) algorithm.

1.5.3: The Third Stage

The role of process variables or work sampling proportions in CLP modeling is then formulated using the collected field data. The fundamental assumption of activity models—that CLP improves if more time is spent on direct work activities—is tested using scatter plots, correlation analysis, and multivariate regression analysis. A linear and non-linear regression analysis is carried out between direct work and output or CLP variables and a multivariate linear regression analysis is carried out between the process and output variables to examine the capability of process variables in providing a credible explanation to the variability of CLP, and based on the results of the null hypothesis tests, inferences on the role of process variables in CLP modeling are made.

Then, a system-based modeling framework to incorporate the key process variables with input variables for improved construction labour productivity modeling and analysis is proposed, developed, and tested. In the system-based modeling framework, depending on the mediation or moderation role of the *P* variables (work sampling proportions) in explaining the variability of CLP, three different paths were considered. The first path is based on the *I–O* relationship and comprises the factor CLP model. The second path is based on the *I–P–O* relationship and assumes that process variables have a mediating effect; it comprises a "mediated system" CLP model. In the mediated system CLP model, the assumption is based on complete mediation, where the *I* variables influence the *P* variables as mediator variables, which in turn influence the output or dependent variable (*O*). The third path is based on the (*I* and *P*)–*O* relationship and assumes that *P* variables have a moderating effect; it comprises a "moderated system" CLP model. In the moderated system CLP model, the assumption is that the *P* variables, as moderator variables, affect the direction and strength of the relationship between the *I* and *O* variables. The mediation and moderation effect of the process variables are tested by developing artificial intelligence technique-based models and evaluating which model and path provided the most accurate results. Timeliness, precision, repeatability, and accuracy performance metrics are used to determine the overall accuracy of a given model path, and the three model paths were tested using field data collected for this research and the most accurate path was identified.

1.5.4: The Fourth Stage

Based on the developed system-based CLP modeling framework, the context-specific and universal CLP models are developed at this stage. By analyzing existing literature in the field of context-aware computing, context and its application in computing fields was examined and, an operational definition of context for CLP modeling is formulated and associated context attributes are developed based on the 5W1H (Who, What, Where, When, Why, and How) question and answers approach. Then, a framework for the development of context-specific CLP models based on fuzzy inference systems (FIS) is developed. The framework first formulates the unique contexts of the studied construction projects using 5W1H approach and projects sharing similar contexts are combined. Additionally a generic CLP model, based on the combined data set of the unique contexts, is developed. Finally, the learning ability of the developed FIS CLP models is improved using a multi-objective optimization framework which optimizes several model parameters for improving the accuracy and interpretability of the developed CLP models.

The context adaptation framework for transferring knowledge among contexts is proposed using linear and non-linear scaling of the membership functions of the context-specific fuzzy CLP models. However, the determination of the parameters of the non-linear function requires an optimization process and genetic algorithm based optimization is used. Both linear and non-linear adaptations are implemented on each of the context-specific CLP models and further sensitivity analysis of the adapted models using fuzzy operators and defuzzification methods is carried out. Then, the performance of the adapted CLP models was evaluated, and the most accurate adaptation technique is identified. Finally, a framework for the development of the universal CLP model is proposed and tested. The framework abstracts the context-specific fuzzy models, which could be used to model and predict CLP for specific contexts, in order to develop a single generalized, more abstract universal CLP model. The development of the universal model is based on a granular fuzzy case-based reasoning approach.

1.6: THESIS ORGANIZATION

This thesis's organization is based on a combination of traditional and paper-based formats, and consists of eight chapters and three appendices. The first two chapters provide the background and research methodology and are followed by five CLP modeling chapters addressing the sequential steps of the CLP model development process. Finally a concluding chapter is presented. Each appendix provides the information associated with the referring chapter.

Chapter 1 provides a brief background of the research, the statement of problems faced in CLP modeling, and the objectives of this research. The expected contributions and a brief methodology of this research are also explained in this chapter.

Chapter 2 reviews existing literature, identifies limitations of past CLP studies, and presents the detailed description of the first stage of the research methodology. The methodology used for identifying, classifying, quantifying, and documenting the influencing input and process parameters together with CLP is described. The results of the extensive data collected for the research is also presented.

Chapter 3 presents the methodology for identifying key input parameters (factors and practices) influencing labour productivity using expert and data-driven approaches. Analysis of the collected factors surveys and results on key parameters influencing labour productivity in building and industrial contexts is presented. Then, the data-driven approach for identifying key parameters using the collected field data is presented and discussed.

Chapter 4 describes the system-based labour productivity modeling framework for establishing the role of work sampling proportions in addition to the input parameters in labour productivity modeling. The formulation and evaluation of the framework is presented and discussed in detail.

Chapter 5 presents the framework for the development of context-specific CLP models based on fuzzy inference systems (FIS). This chapter describes the operational definition of context for CLP modeling, and the procedure for the development, optimization, and validation of a series of context-specific fuzzy inference CLP models.

Chapter 6 presents the context adaptation framework for adapting context-specific models from one context to another. This chapter describes the detailed adaptation process, which is based on linear and non-linear scaling of the membership functions of the context-specific CLP models and further sensitivity analysis of adapted models for fuzzy operators and defuzzification methods, and evaluates the performance of the adapted CLP models.

Chapter 7 presents the framework for the development of universal CLP models. This chapter describes the processes involved in the abstraction of the context-specific models, the development of information granules of the universal model, and the optimization of the universal CLP model.

Chapter 8 describes the conclusions, contributions, and limitations of this research. Recommendations for future research are also included.

1.7: References

Awad, A., and Fayek, A. Robinson. (2013). "Adaptive learning of contractor default prediction model for surety bonding." *J. Constr. Eng. Manage*., 139(6), 694–704.

Botta, A. (2008). “Automatic context adaptation of fuzzy systems.” Ph.D. thesis, IMT Institute for Advanced Studies, Lucca, Italy.

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.

Dai, J., Goodrum, P., Maloney, W., and Srinivasan, C. (2009). "Latent structures of the factors affecting construction labor productivity." *J. Constr. Eng. Manage*., 135(5), 397–406.

Drewin, F.J. (1982). *Construction productivity: measurement and improvement through work study*.1st Ed., Elsevier Science Publishing Inc.

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

Gerek, İ. H., Erdis, E., Mistikoglu, G., and Usmen, M. (2014). “Modelling masonry crew productivity using two artificial neural network techniques.” *J. Civ. Eng. Manage.*, (ahead-of-print), 8pp.

Gouett, M. C., Haas, C. T., Goodrum, P. M., and Caldas, C. H. (2011). “Activity analysis for direct-work rate improvement in construction.” *J. Constr. Eng. Manage*., *137*(12), 1117–1124.

Hanna, A. S., (2010). *Construction labour productivity management and methods improvement*. n.p., Madison, Wisconsin, US.

Harrison, P. (2007). “Can measurement error explain the weakness of productivity growth in the Canadian construction industry?” *Centre for the Study of Living Standards*, Ottawa, Ontario, Canada.

Josephson, P., and Björkman, L. (2013). “Why do work sampling studies in construction? The case of plumbing work in Scandinavia.” *Eng. Construct. Architect. Manage.*, 20(6), 589–603.

Liberda, M., Ruwanpura, J., and Jergeas, G. (2003). "Construction productivity improvement: A study of human, management and external issues." *Proc., Construction Research Congress*, ASCE, Honolulu, Hawaii, 1–8.

Mao, H. (1999). “Estimating labour productivity using fuzzy set theory.” M.Sc. thesis, Univ. of Alberta, Edmonton, Canada.

Moselhi, O., and Khan, Z. (2012). “Significance ranking of parameters impacting construction labor productivity.” *J. Constr. Innov.: Info. Proc. Manage.*, 12(3), 272–296.

Nasir, H., Ahmed, H., Haas, C., and Goodrum, P.M. (2014). “An analysis of construction productivity differences between Canada and the United States.” *Constr. Manage. Econ.,* 32 (6), 595–607.

Oduba, A. O. (2002). “Predicting industrial construction productivity using fuzzy expert systems.” M.Sc. thesis, Univ. of Alberta, Edmonton, Canada.

Oral, M., Oral, E. L., and Aydın, A. (2012). "Supervised vs. unsupervised learning for construction crew productivity prediction." *Autom. Constr.,* 22, 271–276.

Panas, A., and Pantouvakis, J. P. (2010). “Evaluating research methodology in construction productivity studies.” *Built Hum. Env. Rev.,* 3(1), 63–85.

PMI (Project Management Institute). (2007). *Construction extension to the PMBOK (Project management body of knowledge) guide*. Newtown Square, Pennsylvania, US.

Song, L., and AbouRizk, S. M. (2008). "Measuring and modeling labor productivity using historical data." *J. Constr. Eng. Manage*., 134(10), 786–794.

Statistics Canada, (2015a). “Gross domestic product at basic prices, by industry, 2010-2014.” < http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/manuf10-eng.htm> (June 16, 2015).

Statistics Canada, (2015b). “Employment by industry, 2010-2014.” < http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ40-eng.htm> (June 16, 2015).

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). "Modeling construction labor productivity." *J. Constr. Eng. Manage*., 116(4), 705–726.

Tsehayae, A. A., and Fayek, A. Robinson. (2012). “A research framework for work sampling and its application in developing comparative direct and support activity proportions for different trades,” *Proc., Construction Research Congress*, ASCE, West Lafayette, Indiana, US, 1420–1429.

Tsehayae, A. A., and Fayek, A. Robinson. (2014). "Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects." *Can. J. Civ. Eng.,* 41(10), 878–891.

Walker, A. (2015). *Project management in construction.* Blackwell Publishing Ltd, Oxford, UK.

Yi, W., and Chan, A. (2014). "Critical review of labor productivity research in construction journals." *J. Manage. Eng.,* 30(2), 214–225.

Chapter 2: Literature Review and Research Methodology[2]

2.1: INTRODUCTION

In this study, construction labour productivity is modeled using a system approach which involves three model parameters—*Input*, *Process*, and *Output*—so as to examine the cause and effect of the input and process variables on CLP. Construction labour productivity, referred as output parameters (*O*), deals with the efficiency of labour crews converting inputs (land, knowledge, information, energy, materials, etc.) to outputs (project products). Input parameters (*I*) refers to a number of factors and practices (e.g., crew size, crew composition, co-operation among craftsperson, location of work scope, complexity of task, weather condition, risk management practice, etc.), which either directly or indirectly influence CLP. Process parameters (P) refers to work sampling proportions summarizing the actual utilization of labour work hours and provide an in-depth examination of what happens during the conversion process.

In this chapter, first a review of the existing literature on labour productivity models developed to formulate the impact of the influencing parameters on CLP is carried out, limitations are established, and the rational for the proposed system model is presented. Then, the proposed research methodology to fulfil the objectives of this thesis, indicated in the previous chapter, is presented. Additionally, the initial part of the research methodology addressing the identification, classification, quantification, and data collection on model parameters is presented. The identification and classification of influencing input and process parameters, based on review of past labour productivity studies, is presented together with the quantification of the parameters for field data collection. Then, the data collection methodology including the developed research ethics procedure, data collection protocol, and database tool for data storage and analysis is presented. Finally, the results of the extensive data collection process are summarized.

---

[2] Parts of this chapter have been published in Canadian Journal of Civil Engineering, Volume 41, Issue 10, pp. 878-891; the Proceedings, ASCE Construction Research Congress 2014, Atlanta, Georgia, US, May 19-21, pp. 837-846; and submitted for publication in Journal of Construction Innovation: Information, Process, Management, JCI, 36 manuscript pages, submitted July 28, 2015.

2.2: Review of CLP Models

The successful completion of construction projects is highly dependent on construction labour productivity (CLP), which measures the efficiency of construction craftspeople in converting a given set of inputs to tangible outputs. Thus, the ability of estimating team in construction firms to accurately predict CLP values for different activities has a significant impact on the labour cost component of a project, and decision making processes during planning, bidding, and control stages of projects. As a result, numerous predictive modeling approaches as shown in Fig. 2.1 have been developed and tested. Overall, these tested approaches can be categorized as either factor or activity models. Factor models relate the different input variables—made up of key influencing parameters (factors and practices) like crew size, weather condition, etc.—to labour productivity. Activity models mainly relate the process variables, in terms of work sampling proportions, to labour productivity. So far, no study has succeeded in developing an integrated system approach investigating the overall relationship between both input and process variables and CLP. The following subsection reviews and notes the limitations of past studies dealing with factor and activity models.

2.2.1: Factor Models

Several past studies have quantified the impact of different parameters on CLP using factor models. Factor modeling is a multivariate approach to modeling crew-level productivity using influencing parameters (factors and practices) as independent variables and productivity as the dependent variable. Factor model development requires practitioners to gather and measure input parameters, and identify the key independent (i.e., input) variables and to model the complex relationship between these variables and productivity (the output) using appropriate analysis methods. Accordingly, numerous parameters that either directly or indirectly influence CLP have been presented, and critical parameters or variables were usually identified according to the study objective, and were then documented and applied in CLP modeling. Such studies have often relied on author's knowledge and factor surveys with groups of experts to establish critical parameters, before proceeding with data collection. Alternatively, when detailed parameter documentation is carried out, data-driven analysis methods can used to identify the critical parameters for modeling CLP.

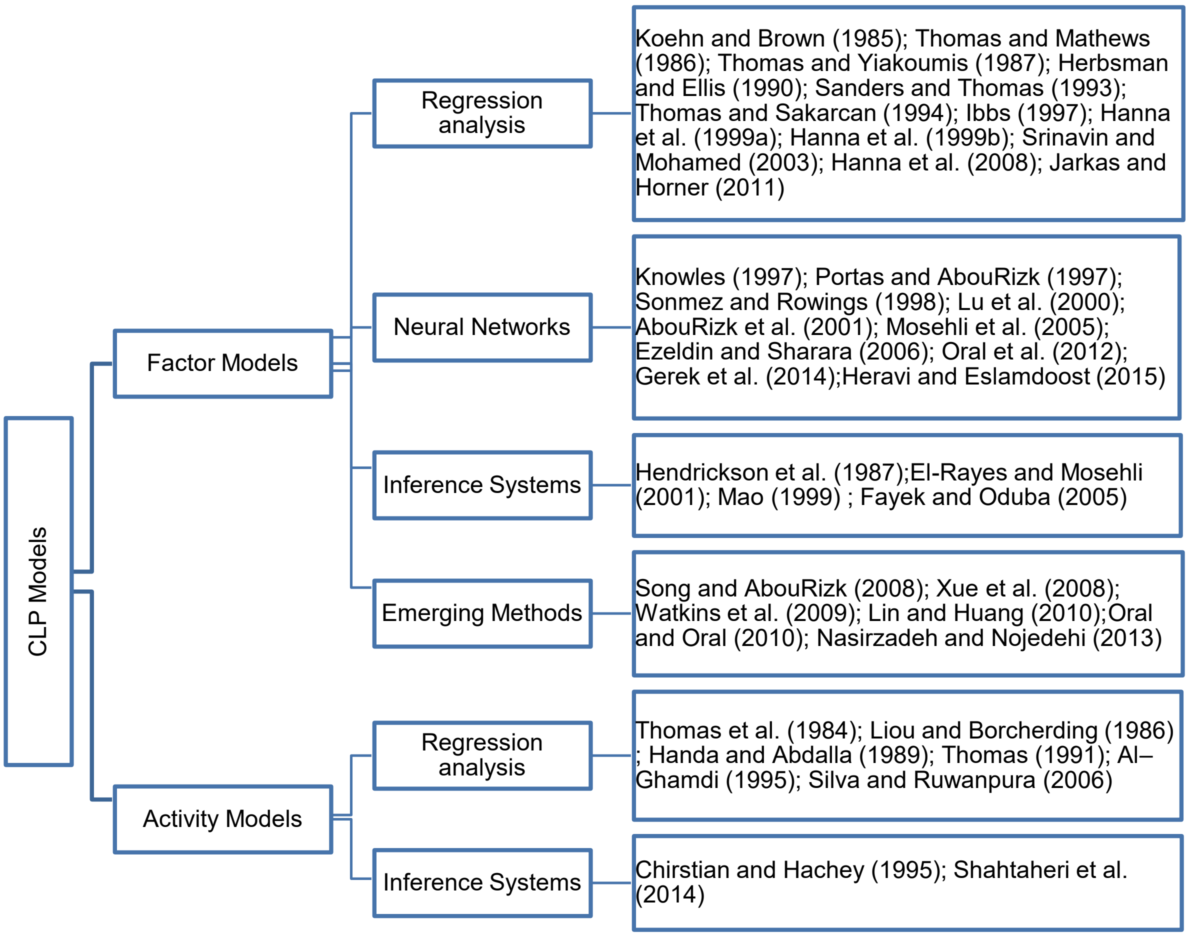

Figure 2.1: Summary of Past CLP Models

Numerous factor modeling approaches for CLP have been developed based on different data analysis methods including: regression analysis, artificial neural networks, expert systems, fuzzy inference systems, discrete event simulation, system dynamics, and agent based simulation. The following subsections present a review of the different factor models, which have been grouped into regression analysis, artificial neural networks, inference systems, and other emerging methods, and limitations are summarized.

2.2.1.1: Regression Analysis

The relationship between the assorted input variables (key parameters) and CLP, and the degree of their impact on it, has most frequently been modeled using regression analysis. The method relies on actual data and curve fitting techniques between a set of independent variables and productivity as a dependent variable. Once the regression models are developed, normality of the residuals (predicted minus observed values) has to be verified using standardized histograms, normality plots, and scatter plots of residuals so as to ensure the validity of the developed modes. In a multivariate factor model for modeling labour intensive crew-level productivity, it was argued that although factors affecting crew level productivity can lead to random or systematic, controlled or uncontrolled disturbances, they can be individually isolated to quantify the effects of each factor on ideal productivity (Thomas and Yiakoumis 1987). Subsequently, most existing factor models address the effect of a single input variable, like temperature, on CLP (Yi and Chan 2014).

Koehn and Brown (1985) studied the relationship between productivity and weather factors (temperature and humidity) by combining data from different activities (manual excavation, erection, masonry, electrical, carpentry, labourer, and equipment excavation), and provided two non-linear equations for cold and warm weather conditions. Thomas and Mathews (1986) studied five learning curve models using precast concrete floor planks erecting activity data, and concluded that the cubic learning model is the best predictor of productivity unit rates. Thomas and Yiakoumis (1987) studied masonry, structural steel, and formwork activities and using multiple regression analysis, developed a factor model for predicting the performance factor, defined as a ratio of actual to expected productivity. Herbsman and Ellis (1990) using regression analysis, investigated the relationship between influencing factors and productivity of steel form erection activity. Moselhi et al. (1991) studied the effect of change orders on productivity of civil, electrical, and mechanical works. Sanders and Thomas (1993) developed a statistical model to forecast the productivity of masonry activities based on an additive regression model and data collected from 11 masonry projects. Thomas and Sakarcan (1994) developed a factor model for masonry projects as shown in Eq. (2.1), where E_t = predicted productivity for time period t; I_s = productivity for standard conditions; m = number of condition variables; a_i = coefficient of condition variable *i*; x_i = indicator of condition variable *i* (0 = not present, 1 = present); n = number of submodels; and $f(y)_j$ = mathematical function of submodel *j*.

$$E_t = I_s + \sum_{i=1}^{m} a_i \, x_i + \sum_{j=1}^{n} f(y)_j \tag{2.1}$$

Ibbs (1997) studied the impact of change orders on productivity during design and construction phases using regression. The impact of change orders on labour efficiency was also investigated for electrical and mechanical work (Hanna et al. 1999a; Hanna et al. 1999b). Srinavin and Mohamed (2003) developed a thermal factor model using air temperature, relative humidity, radiant temperature, wind velocity, and nature of task, and indicated that the developed model performed satisfactorily in light and moderate tasks, but much less in for heavy tasks. Hanna et al. (2008) assessed effects of employing additional shift work during acceleration of construction schedule, and showed that shift work has the potential to be both beneficial and detrimental to CLP. Jarkas and Horner (2011) revisited the applicability of learning curve theory in CLP analysis and showed that there was no significant improvement in formwork labour productivity due to learning.

However, the above factor models based on regression method have a number of major limitations including: lack of capacity to deal with numerous factors, intolerance to noisy data, and impracticality of deciding the best fitting curve (linear, quadratic, etc.) for representing the highly complex nonlinear relation between the input variables and CLP (Thomas et al. 1990; Lu 2001). Additionally, multiple regression analysis requires each input variable to have a linear relationship with CLP, input variables not be correlated with one another, and residuals (predicted minus observed values) to be normally distributed. As, the input variables (factors) in CLP analysis are often related to one another (Nasirzadeh and Nojedehi 2013), and the relationship between input variables and CLP is highly complex and nonlinear, the stated requirements of multiple regression will be violated in modeling CLP, and the produced models will produce inefficient predictions.

2.2.1.2: Neural Networks

More recent CLP studies focus on the use of artificial neural networks (NNs). Knowles (1997) presented a two-stage neural network model for predicting CLP for concrete formwork activity in commercial walls and slabs, and industrial pipe handling and welding activities. Portas and AbouRizk (1997) developed a three-layered feed-forward, back-propagation NN model with fuzzy output layer to estimate the likelihood of the production rate of concrete formwork activity. Sonmez and Rowings (1998) developed labour productivity models for concrete pouring, formwork, and concrete finishing activities using neural networks, and showed that NNs have the potential for quantitative evaluation of the nonlinear effect of multiple interacting factors on CLP. Lu et al. (2000) used probability inference neural networks to develop CLP models for field pipe installation and shop spool fabrication activities. AbouRizk et al. (2001) developed a two-staged neural network based CLP model for industrial welding and pipe installation activities. Mosehli et al. (2005) utilized NN for modeling the impact of change orders on labour-intensive operations based on historical company-specific data. Ezeldin and Sharara (2006) developed CLP prediction models for forms assembly, steel fixing, and concrete pouring activities using feed-forward back-propagation neural networks, and indicated that the concrete pouring CLP model was the least accurate one as compared to others. Oral et al. (2012) compared the performance of feed-forward neural networks, generalized regression neural networks, and self-organizing maps in predicting construction crew productivity for plastering crews and indicated that the self-organizing maps had better prediction ability. Gerek et al. (2014) also compared the performance of feed-forward neural networks and radial basis neural network in modelling the productivity of masonry crews, and showed that radial basis NN performed better, although both slightly overestimated the masons' productivity. Heravi and Eslamdoost (2015) studied labour productivity of concreting work for gas, steam, and combined cycle power plant construction projects using neural networks.

Neural networks provide an effective tool for complex problems, such as modeling CLP where the relationships between inputs and output cannot be easily represented by mathematical functions (Moselhi et al. 1991). However there are limitations to its application for construction labour productivity studies.

NN models will first need high quality data, which is difficult to guarantee in CLP studies, as poor or insufficient data could result in an incorrect CLP model (Huh 2004). Additionally, the mapping of the input and output data is not interpretable or transparent and makes NNs difficult to understand, thus, limit the application of NN models. Finally, neural network models are not appropriate for adaptation to suit other contexts, as users cannot calibrate the developed model without going through the process of retraining.

2.2.1.3: Inference Systems

Because CLP modeling is a complex problem with limited data availability, and deals with a large number of subjective variables, CLP modeling is an exceptional target for another artificial intelligence technique: inference systems. Two types of inference systems are recognized: expert systems and fuzzy inference systems. While fuzzy inference systems are based on fuzzy logic and if-then rules; expert systems are based on traditional two valued logic systems. Expert systems were first used in construction productivity to predict activity duration and productivity for masonry construction (Hendrickson et al. 1987). The expert system, called MASON, was developed based on interviews with one professional mason and one supporting labourer, but the model was not validated, and the system relies on two staged approach of predicting the maximum expected productivity followed by experts based adjustments to establish realistic productivity estimates (Lu 2001). El-Rayes and Mosehli (2001) also created a database of climatic historical data and combined it with knowledge-based rules to create an expert system, called WEATHER that could estimate the lost productivity due to rainfall on highway construction.

Construction researchers have found the use of fuzzy logic useful due to its focus on human thinking and natural language than the traditional two valued logic systems (Mao 1999). However, there are few applications of fuzzy logic in the CLP modeling field (Yi and Chan 2014). Mao (1999) used fuzzy inference system to model a labour productivity of concrete wall formwork activity, and used an assumed membership functions together with a rule-base developed based on historical data; however, the data used did not cover all input factors. Fayek and Oduba (2005) also used fuzzy inference systems to model industrial pipe rigging and welding activities CLP, the models showed high linguistic accuracy but the numeric accuracy was low.

However, several limitations in inference systems are also observed. Expert systems are mainly advantageous in very narrow defined problems; additionally, experts used to develop the systems might lack the ability to map the inputs with output, and also expert systems cannot be adapted to changing environments (Lu 2001). Fuzzy inference systems (FISs) have proved effective tools for solving engineering problems in biomedicine, robotics, pattern recognition, image processing, and control application areas, and are suitable for adaption to suit other environments (Botta 2008). Thus, in this research the developed labour productivity models are based on fuzzy inference systems. However, there have been few applications of FISs in CLP modeling, and the few studies using FISs were limited in that they did not develop membership functions and if-then rules using adequate data (Mao 1999; Fayek and Oduba 2005). Additionally, FIS inability to learn from data and develop and optimize system parameters is a major limitation that needs to be addressed in CLP modes.

2.2.1.4: Emerging Methods

Recently a number of advanced methods are being employed in modeling CLP. Song and AbouRizk (2008) studied steel fabrication and steel drafting activities and combined discrete-event simulation with a neural network to model the productivity of a production system that had a number of related activities. The NN was used to model individual activities and the complex relationship between productivity and influencing factors, and discrete-event simulation was used to simulate the entire shop fabrication production process.

Data envelopment analysis (DEA) measures the relative efficiency of decision-making units, without specifying a function to express production relationship between inputs and outputs. Xue et al. (2008) used DEA based Malmquist productivity index to measure the productivity change over time in Chinese construction industry. Lin and Huang (2010) also applied DEA for deriving baseline productivity and concluded that DEA is the best method to derive contractors' relative performances and benchmarks for best practice. Oral and Oral (2010) developed CLP models for concrete pouring, formwork, and reinforcement placement activities using self-organizing maps. Nasirzadeh and Nojedehi (2013) studied the highly dynamic nature of the influencing factors throughout the life cycle of the project using system dynamics technique, and the complex inter-related structure of different factors was modeled using cause and effect feedback loops and a qualitative CLP model was developed for concrete pouring activity.

Agent based modeling (ABM) is a computer simulation technique that allows the examination of how system rules and patterns emerge from the behaviors of individual agents. ABM creates artificial agents that represent individuals that have the ability to perceive and interact with each other and their environment and based on their interactions make autonomous decisions (Archiszewski et al. 2005). Watkins et al. (2009) used ABM to represent a construction site as a system of complex interactions and explored whether labour efficiency can be treated as an emergent property resulting from individual and crew interactions in space; thereby allowing for a “bottom-up” approach to analyzing labour efficiency.

2.2.2: Activity Models

Activity models relate labour time utilization measures like work sampling proportions (i.e., process parameters) to CLP. How construction workers spend their working time is of great concern to construction companies. Under the influence of lean thinking in the construction industry, interest in labour time utilization and eliminating waste is rising (Yi and Chan 2014). Work sampling (WS), a widely used work-study method in the construction industry, uses random observation to investigate how a workforce uses its work time. WS establishes the percentage of work time spent on categories, and direct work (tool time), which represents the proportion of work time spent exerting physical effort directed toward the completion of an activity, has been used as a surrogate measure of CLP (Thomas et al. 1991; Gouett et al. 2011). However, WS studies have been inconsistently implemented (Yi and Chan 2014). The activity model is based on WS and is readily applicable to labour-intensive activities. A valid activity model is required to show that direct work times and outputs are related in some predictable fashion (Thomas et al. 1991; Yi and Chan 2014). However, past studies have shown that the definition of work categories and the subsequent task classifications can significantly affect the different proportions, and, hence, their relationships with CLP (Thomas et al. 1991).

The relationship between process variables and CLP has so far been investigated using mainly regression analysis, while some studies have used inference systems. There are two opposing views of the validity of activity models. For the most part, the literature argues that WS can be used to predict productivity (Josephson and Björkman 2013). Thomas et al. (1984) studied the relationship between direct work percentage and productivity, defined as a ratio of earned man-hours to actual man-hours, and established an activity model for pipefitting activity. The study reported a positive Pearson correlation coefficient of 0.86 between productivity and direct work category. As is common in any regression analysis, the dependent and independent variables have to be carefully formulated. Contrary to Drewin's assumption that labour productivity is the dependent variable and WS data (direct, delay, etc.) is the independent variable (1982), Thomas et al. (1984) assumed crew output as independent and direct work values as dependent. Liou and Borcherding (1986) statistically proved that WS results strongly correlate with unit-rate productivity for power plant construction projects. By using 41 data points for concrete work elements, the authors demonstrated a relationship between unit productivity (concrete man hours/cubic yard) and different WS activity categories (direct work, material or equipment handling, late and break). Similar to Thomas et al. (1984), unit productivity was the independent variable and the WS components were assumed to be the dependent variable. A detailed study on Canadian housing sector showed that process variables could be used to indicate actual site productivity and crew learning rates (Handa and Abdalla 1989). A study on concreting operations of four commercial construction sites in Alberta, Canada, also developed an activity model, with a Pearson correlation coefficient of 0.90 between productivity (m^3/hr) and tool ratio or direct work proportion (Silva and Ruwanpura 2006).

Chirstian and Hachey (1995) developed a prediction model for production rates of concrete placement activities using an expert system. The model was developed based on concrete placement data from several projects, and relied on question and answer routine to develop a domain expert rule-base; however, there was significant variation in data sources, and sample size was limited and the developed rule-base was inconsistent (Sonmez 1996; Lu 2001). Shahtaheri et al. (2014) assumed that labour performance can be improved by increasing the direct-work rate, and developed a fuzzy inference model to estimate appropriate baselines and set realistic goals for direct-work rate.

Conversely, Thomas (1991) stated that direct work is not related to productivity; using data from seven databases (five papers and two data sets) containing over 158 WS studies, mainly from nuclear power plant projects, he concluded that previous studies lacked validity. An activity model study on wall building activities using work sampling observations also concluded that work sampling is not a strong predicator of productivity (Al–Ghamdi 1995). According to the author, the study which used work sampling data of six masons in 35 field experiments showed a Person's correlation of coefficient of 0.498, -0.675, and -0.914, between direct or effective work categories and productivity, between support or essential contributory work categories and productivity, and delay or ineffective work categories, respectively.

2.2.3: Limitations of Past CLP Models

Despite its obvious importance in construction project management, developing accurate and interpretable CLP models for analysis and improvement of construction productivity has not been fully achieved (Yi and Chan 2014). Modeling CLP remains a challenge due the complex variability of CLP, the limited data availability to study CLP under various contexts, and the requirement of considering the complex impact of the multiple subjective and objective variables simultaneously, while maintaining a high accuracy and interpretability in the developed models. However, most of the past CLP studies focused on limited number influencing variables and additional researches are required for the identification of the multilevel, complex, and context-dependent key variables influencing CLP in different project contexts. The factor and activity models discussed above were developed for specific contexts and their implementations were mostly restricted to the information used to develop each specific model. However, only a few CLP studies had an explicit definition of the context of the CLP modeling processes, and overlooked the role of context in CLP modeling and its importance for formulating the circumstances that form the setting of the CLP model and its development process.

Most of the factor models were not able to deal with subjective variables in a comprehensive manner, and due to the high dimension and non-linear nature of the CLP modeling problem, recent CLP studies are focusing on neural networks, which are not interpretable by users. Furthermore, activity models lacked statistically significant results, and are limited by their inability to model the effect of the process parameters in influencing CLP and their dependence on assumptions regarding category definitions, caused due to a lack of standardization (Josephson and Björkman 2013). Additionally, in past CLP studies a lack of an integrated system approach, that investigates the overall relationship between both input and process variables and CLP, is observed. As a result of these limitations, the above CLP models have been unable to provide useable solutions for the highly complex, context-dependent, and non-linear modeling problem of CLP.

Construction productivity is one of the most studied areas in construction engineering and management field (Yi and Chan 2014). As discussed above, numerous predictive CLP models have been developed. However, in CLP modeling field an approach for transferring the knowledge represented in such models from one context to another context is also lacking. Such an approach is particularly important when modeling new contexts for which data availability is limited; and existing models cannot be applied without some adaptation. Additionally in CLP modeling, a universal or generalized modelling framework to abstract the various CLP models and represent a versatile knowledge is also missing. Such generalized CLP models are useful for analyzing CLP irrespective of the context of the projects.

2.3: RESEARCH METHODOLOGY

It is well known that research yields unique and significant results improving the body of knowledge, and innovative research is vital in today's highly complex construction process (Halpin 2007). With the increasing complexity of construction projects and their performance requirements, the role and expectations of construction management practices and research endeavors are becoming even more critical. Proven management practices are often based on relevant and successful past experiences which have been validated by sound research methods to develop validity and reliability (Richard and Liu 2008). In this research, construction labour productivity is considered as an applied research involving the study of an open-ended system, as shown in Fig. 1.1. Additionally, this study mainly focuses on inductive or problem solving research approach and uses a mix of qualitative (factor surveys, foreman delay surveys, and craftsman questionnaires) and quantitative (factors and practices documentation, work sampling observations, and productivity measurement) strategies to establish the methodology for developing a series of context-specific and universal CLP models. The research methodology of this research, developed to fulfil the research objectives described in Section 1.3, is diagrammatically explained in Fig. 2.2.

In the following subsections the initial part of the research methodology dealing with the identification and quantification of the system parameters is described in detail. Additionally the data collection methods and data collections results are discussed. Chapters 3—7 discuss the CLP modeling steps 1—5 (see Fig. 2.2) and present the respective research methodologies together with the data analysis and modeling approaches specific to the part of the research presented in the particular chapter.

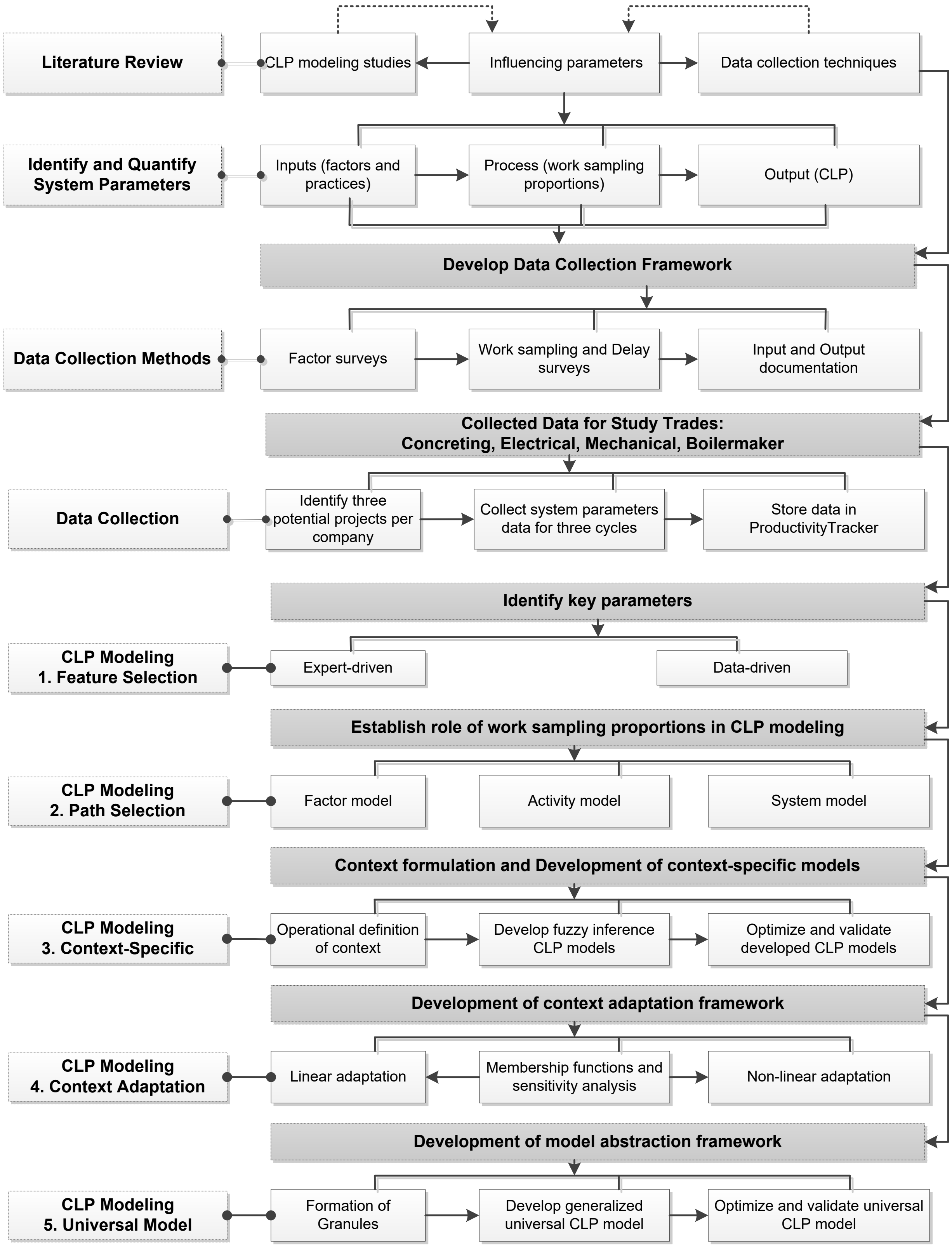

Figure 2.2: Research Methodology

2.3.1: Identification and Quantification of Input Parameters Influencing CLP

Because of its significance to project performance, CLP has been well studied, and numerous parameters that influence CLP have been identified. Past studies generally grouped parameters across internal and external categories, and varying number of parameters were established. Lim and Alum (1995) identified 17 influencing factors grouped under three categories—manpower, management, and environment. Oduba (2002) identified 57 influencing factors grouped under context and input; the factors were further classed according to project or activity level. Liberda et al. (2003) identified 51 influencing factors grouped under three categories—human, external, and management. CII (2006) identified 83 influencing factors grouped under eleven categories: supervisor direction, communication, safety, tools and consumables, materials, engineering drawing management, labour, foreman, superintendent, project management, and construction equipment. Durdyev and Mbachu (2011) identified 56 influencing factors grouped under two main categories—internal constrains and external constraints. Jarkas and Bitar (2012) identified 45 influencing factors grouped under four categories—technological, human/labour, management, and external. The summary of the identified parameters, presented in Appendix A.1, shows that despite the extensive research in the area, consensus on the classification of parameters, listings of the context-specific parameters, and generalization of common parameters is yet to be achieved. Furthermore, past CLP modeling studies did not investigate project practices as key factors influencing CLP (Tsehayae and Fayek 2014a).

In this study, input parameters, referred as I, and made up of factors (all factors influencing CLP except for project practices) and practices (practices related to the management of the project based on Project Management Institute [PMI] 2007, and identified best practices based on Construction Industry Institute [CII] 2011) were investigated. A total of 169 input parameters influencing CLP were thus gathered from related literature (CII 2006; Liberda et al. 2003; Jergeas 2009; Knight and Fayek 2000; Oduba 2002; O'Connor and Huh 2006; AbouRizk et al. 2001; Chan et al. 2004; Dissanayake et al. 2005). Construction practices like the use of a material tracking system have been shown to significantly improve CLP (Grau et al. 2009; Chanmeka et al. 2012); however, such practices have not been properly integrated into CLP analysis and modeling efforts. The developed list of parameters was verified using factor surveys with respondents from project management and trade groups.

Labour productivity naturally tends to be a micro level issue where a group of organized workers are required to transform a set of inputs to tangible project outputs (Drewin 1982). However, the factors and practices directly or indirectly influencing labour productivity are multileveled ranging from macro to meso and then to micro levels (CII 2006; Knight and Fayek 2000). In order to capture and demonstrate their multilevel nature, a hierarchal structure was adopted wherein the input parameters were classified according to the following six levels: activity, project, organizational, provincial, national, and global. The activity level parameters were further grouped under labour and crew, materials and consumables, equipment and tools, task property, location property, foreman, and engineering and instructions categories. The project level parameters were further grouped under project delivery/design document, project nature, project condition, project owner, project team, project labour and union, salary and benefits, and project practices (initial planning, scope, time, cost, quality, procurement, safety, risk, communication, human resource, environmental, and claim management) categories. In Table 2.1, the full list of the input parameters is presented, which introduces a number of factors and practices that were not included in previous research works, and provides a solid foundation that researchers and industry practitioners can use to conduct a complete study on the causes of CLP variation among projects, so as to obtain more useful insights into which factors and practices lead to improved CLP. In Table 2.1, the identification labels of each input parameter are also included.

Table 2.1: Input Parameters (Factors and Practices) Influencing CLP

Parameter category	**Activity level parameters, ID**
Labour and crew	Crew Properties [Crew size$_{x1}$, Craftsperson education$_{x2}$, Craftsperson on job training$_{x3}$, Craftsperson technical training$_{x4}$, Crew composition$_{x5}$, Crew experience (seniority)$_{x6}$, Number of languages spoken$_{x7}$, Co-operation among craftsperson$_{x8}$], Craftsperson learning effect$_{x9}$, Treatment of craftsperson by foreman$_{x10}$, Craftsperson motivation$_{x11}$, Craftsperson fatigue$_{x12}$, Craftsperson trust in foreman$_{x13}$, Team spirit of crew$_{x14}$, Level of absenteeism$_{x15}$, Crew turnover$_{x16}$, Discontinuity in crew makeup$_{x17}$, Level of interruption and disruption$_{x18}$, Fairness of work assignment$_{x19}$, Crew participation in foreman decision-making process$_{x20}$, Crew flexibility$_{x21}$, Job site orientation program$_{x22}$, Job security$_{x23}$, Availability of craftsperson$_{x24}$, Multiskilling of crew$_{x25}$
Materials and consumables	Task materials: availability $_{x26}$ and quality$_{x27}$, Temporary material storages (availability, distance, travel time)$_{x28}$, Consumables (availability, policy)$_{x29}$, Material tracking system$_{x30}$, Material unloading practices$_{x31}$, Material movement practices: horizontal $_{x32}$ and vertical$_{x33}$
Equipment and tools	Availability of work equipment$_{x34}$, Availability of transport equipment$_{x35}$, Equipment breakdown$_{x36}$, Availability of tools $_{x37}$, Tools management (sharing, quality, location and efficiency of tool room attendant, misplacement, quality of maintenance)$_{x38}$, Availability of electric power$_{x39}$, Availability of extension cords$_{x40}$
Task property	Task (complexity, repetitiveness, total work volume)$_{x41}$, Rework: level $_{x42}$ and frequency$_{x43}$, Task change orders$_{x44}$, Placement technique$_{x45}$, Building/structural element$_{x46}$
Location property	Working condition (noise)$_{x47}$, Location of work scope: distance$_{x48}$ and elevation$_{x49}$, Congestion of work area$_{x50}$, Cleanliness of work area$_{x51}$, Cover from weather effect$_{x52}$, Location of facilities (lunch rooms, washrooms)$_{x53}$
Foreman	Foreman experience$_{x54}$, Training$_{x55}$, Skill and responsibility$_{x56}$, Fairness in performance review of crew by foreman$_{x57}$, Change of foremen$_{x58}$, Span of control$_{x59}$, Use of assistant foremen$_{x60}$, Provision of feedback on foreman's performance$_{x61}$
Engineering and instructions	Drawings (availability, quality, number of revisions) $_{x62}$, Specifications (use of standard specifications, availability, quality of specification)$_{x63}$, Response rate with request for information affecting task at hand$_{x64}$, Adequacy of instructions$_{x65}$
Parameter category	**Project level parameters, ID**
Project delivery/Design document	Project delivery (delivery system, contract type, level of fast tracking)$_{x66}$, Changes (design drawings, specifications, contract conditions)$_{x67}$, Lack of information$_{x68}$, Approval for building permit$_{x69}$
Project nature	Type$_{x70}$, Size $_{x71}$, Complexity$_{x72}$, Location$_{x73}$, Year of construction$_{x74}$, Amount of modularization$_{x75}$, Organizational structure$_{x76}$, Project level rework$_{x77}$, Project level change order$_{x78}$, Project percent complete$_{x79}$
Project condition	Site transportation (flight arrangements, provision of ground transportation for workers to site)$_{x80}$, Camp condition$_{x81}$, Weather: Temperature$_{x82}$, Precipitation$_{x83}$, Humidity$_{x84}$, Wind speed$_{x85}$, Radiation$_{x86}$, Variability of weather (number of heating days, cold days)$_{x87}$, Ground conditions$_{x88}$, Site layout$_{x89}$, Site congestion$_{x90}$, Site access (queue time to access site)$_{x91}$, Parking facilities (within project)$_{x92}$, Site main storage$_{x93}$, Site facilities for workers (lunch room, wash room)$_{x94}$, Unloading/laydown area$_{x95}$, Project work times (use of overtime, multiple shifts, shift length)$_{x96}$

Table 2.1: Input Parameters (Factors and Practices) Influencing CLP (continued)

Parameter category	**Project level parameters, ID, continued**
Project owner	Owner staff on site$_{x97}$, Supervision from owner or representative$_{x98}$, Owner's primary driver$_{x99}$, Delivery of site to contactor$_{x100}$, Approval of shop drawings and sample materials$_{x101}$, Suspension of project work (owner reasons)$_{x102}$
Project team: Project manager, superintendent	Experience of project management team members$_{x103}$, Support and administrative staff$_{x104}$, Level of paper work for work approval$_{x105}$, Treatment of foremen by superintendent and project manager$_{x106}$, Performance competition system within the company$_{x107}$, Uniformity of work rules by superintendent$_{x108}$, Superintendent (education, training)$_{x109}$, Project Manager (education, training)$_{x110}$
Project labour and union	Labour union (type and influence)$_{x111}$, Availability of labour$_{x112}$, Labour disputes (legal cases between a worker on a project)$_{x113}$
Salary and benefits	Salary (project manager, superintendent, foreman, craftsperson: journeyman and apprentice)$_{x114}$, Benefits$_{x115}$
Project initial planning practice	Detailed front end planning$_{x116}$, Constructability review$_{x117}$
Project scope management practice	Project scope (definition, verification, change control)$_{x118}$
Project time management practice	Project planning (activity definition, activity sequencing, activity duration)$_{x119}$, Project scheduling (project duration, criticality of project schedule)$_{x120}$, Project schedule control (schedule compression, activity weights definition, project progress curves development and progress monitoring)$_{x121}$
Project cost management practice	Project resource planning$_{x122}$, Project cost estimating (development of material and equipment list, estimation team experience, time allowed for estimation, bidding climate, labour climate)$_{x123}$, Project cost budgeting$_{x124}$, Project cost control$_{x125}$, Labour productivity measurement practice$_{x126}$
Project quality management practice	Project quality planning$_{x127}$, Project quality assurance$_{x128}$, Project quality control (out of sequence inspection or survey work)$_{x129}$
Project procurement management practice	Procurement (planning, solicitation)$_{x130}$, Procurement administration (material, equipment, tool)$_{x131}$, Trade subcontracting (Subcontracted amount, Number of subcontractors)$_{x132}$
Project safety management practice	Project safety planning$_{x133}$, Project safety plan execution (use of daily job hazard assessment forms, use of site safety meetings, construction equipment safety procedure, drug testing, safety inspections, safety audits, adequacy of protective gear, uniformity of safety procedures)$_{x134}$, Safety training $_{x135}$, Safety incidents$_{x136}$, Safety incident investigation $_{x137}$, Project safety administration and reporting$_{x138}$
Project risk management practice	Risk planning (identification, planning, use of risk assessment tool)$_{x139}$, Risk monitoring and control$_{x140}$, Crisis management$_{x141}$
Project communication management practice	Project communication (communication between different trades)$_{x142}$, Availability of communication devices)$_{x143}$
Project human resource management practice	Project interface development$_{x144}$, Project team (team development)$_{x145}$

Table 2.1: Input Parameters (Factors and Practices) Influencing CLP (continued)

Parameter category	**Project level parameters, ID, continued**
Project environmental management practice	Environmental rating of project$_{x146}$, Project environmental assurance (sorting of waste material)$_{x147}$
Project claim management practice	Project claim (identification, quantification)$_{x148}$, Project claim resolution (resolution method, resolution process)$_{x149}$
Parameter category	**Organization level parameters, ID**
Organizational	Organization's nature (principal project type, year in industry, annual turnover in dollars, structure, work execution approach)$_{x150}$, Annual employee turnover$_{x151}$, Project load$_{x152}$
Parameter category	**Provincial level parameters, ID**
Provincial	Provincial (economy, income tax, GST, unemployment rate of construction workers, provincial codes and regulations,)$_{x153}$, Labour strikes$_{x154}$, Total number of project within province$_{x155}$, Available supervisor pool in province$_{x156}$, Construction material fluctuation$_{x157}$, Expenditure level towards projects (residential, non-residential, energy)$_{x158}$, Cost of project (index)$_{x159}$
Parameter category	**National level parameters, ID**
National	Political system$_{x160}$, Competing project across the nation$_{x161}$, National labour (availability of labour in nation, foreign workers recruitment)$_{x162}$, Canada population (size of population, growth of population, aging of population)$_{x163}$, Economy (interest rate, inflation rate)$_{x164}$
Parameter category	**Global level parameters, ID**
Global	Global economy (outlook, energy demand and supply)$_{x165}$, Oil price$_{x166}$, Oil price fluctuation$_{x167}$, Natural gas price$_{x168}$, Natural gas fluctuation$_{x169}$

The vital and starting point of any CLP analysis and modeling study involves the quantification of input parameters (factors and practices) influencing CLP. However, the parameters affecting labour productivity are numerous, complex, interlinked, and dynamic thus making quantification and data collection a challenging task. Additionally, quantification of the parameters is complicated as the factors and practices are a mix of subjective and objective concepts and require the development of an appropriate measurement scheme (Thomas et al. 1990). Parameters having subjective concepts like fairness of foreman in work assignment or uniformity of safety procedures require detailing of the parameter to a level that accurate data can be collected. Though measurement of objective parameters, such as temperature and crew size, has been easy to carry out, measurement of subjective parameters like supervision skill of superintendent has presented challenges that researchers have attempted to address through the use of simple rating scales without calibration of each measurement scale (Oduba 2002; Thomas et al. 1990). As a result, past CLP studies have tried to first identify the critical parameters based on expert knowledge before completing detailed measurements so as to simplify the data collection process (Thomas et al. 1990; AbouRizk et al. 2001; Chan et al. 2004; Dai et al. 2009). Unfortunately, this deductive approach has not improved understanding of the parameters and their impact on the complex construction process (Panas and Pantouvakis 2010).

Additionally, in rare cases where detailed parameter quantification and documentation was carried out together with documentation of the output parameter (achieved labour productivity), data-driven method could be employed to identify critical parameters; as data-driven methods like correlation analysis, feature reduction, and principal component analysis have been useful in identifying critical parameters, resulting in better prediction ability (Gray and MacDonell 1997; Jang et al. 2011; Moselhi and Khan 2012). Unfortunately, quantifying and documenting the number of parameters known to affect labour productivity is not an easy task so it has rarely been tried let alone achieved to a level at which the actual parameters could be determined using data-driven techniques (Moselhi and Khan 2012).

Therefore, in this research, for each of the 169 input parameters identified from existing literature, a measurement scale must be developed so as to quantify the input parameters and enable construction site data collection. The input parameters were first verified with experts from different levels of the project management teams; the detail discussion of the process is presented in Chapter 3. Quantification of the numerous input parameters influencing CLP is itself a topic of research, so past researchers have opted to measure only a select few (Yi and Chan 2014). This research uses data collected on all 169 identified input parameters to provide complete coverage of all issues influencing CLP. Naturally, some of these parameters are objective—with explicit concepts like *crew size*—while others are subjective—with implicit concepts like *fairness of work assignment*—so the need to develop appropriate measurement schemes for each kind of parameter compounds the challenge presented by their numerousness. As, parameters with subjective concepts must be defined in detail so their measurement scales can function across different project contexts, this study developed a measurement scheme for all identified parameters using appropriate objective and subjective measurement scales. Table 2.2 shows some example measurement scales for objective (e.g., *weather (wind speed)*) and subjective (e.g., *fairness of work assignment* and *foreman skill in resource allocation*) parameters.

Objective parameters have well-defined numerical measures (e.g., *crew size* is measured in terms of number of workers). Subjective parameters (e.g., foreman skill in resource allocation) lack well-defined measurement schemes; hence, for each of them a pre-determined 1–5 scale based on sub-parameter has been developed. The use of sub-parameter is required when a parameter cannot be measured directly and an indicator is needed that suggests the extent of the parameter's existence. Sub-parameters are based on explicit concepts associated with the parameter. For example, the *fairness of work assignment* parameter's sub-parameters were defined based on consistency, reasonableness, and information provision (Sheppard and Lewicki 1987). The sub-parameters enabled the development a predetermined 1–5 rating scale capable of measuring the parameter (see Table 2.2). By developing descriptions of what each predetermined rating scale represents, this study standardizes the documentation of subjective parameters so they remain relevant no matter the project (Awad and Fayek 2012). The quantification process to formulate appropriate measurement scales and facilitate the documentation of the input parameters (factors and practices) influencing CLP resulted in a total of 314 detailed sub-parameters. Accordingly, there were 96 sub-parameters at activity level, 180 at project level, 7 at organizational level, 15 at provincial level, 9 at national level, and 7 at global level.

Table 2.2: Examples of Input Parameter Quantification

Parameter	Scale of measure (unit)	Cycle	Data source	Sample data
Crew size	Integer (total number of crew members)	Daily	Researcher	6
Crew composition	Proportion (ratio of journeymen to apprentices)	Daily	Foreman	0.5
Fairness of work assignment	1–5 Predetermined rating [a]	Daily	Crew members	3
Foreman skill in resource allocation	1–5 Predetermined rating	Daily	Superintendent	4
Number of drawing revisions	Integer (number of drawing revisions per week)	Weekly	Project Manager	0
Change of foremen	Turnover rate (turnovers per month)	Monthly	Superintendent	1
Project location	Categorical	Initially	Researcher	Nisku

Note: [a] Descriptions for predetermined scale ratings are: (1) INCONSISTENT work assignment on daily basis, UNREASONABLE work assignment among crew members, VERY POOR information provision; (2) INCONSISTENT work assignment on a daily basis, UNREASONABLE work assignment among crew members, POOR information provision; (3) SOMEWHAT CONSISTENT work assignment on a daily basis, REASONABLE work assignment among crew members, AVERAGE information provision; (4) VERY CONSISTENT work assignment on a daily basis, REASONABLE work assignment among crew members, GOOD information provision; and (5) VERY CONSISTENT work assignment on a daily basis, REASONABLE work assignment among crew members, VERY GOOD information provision.

Next, the appropriate data collection cycle and data source for each sub-parameter, shown in Table 2.2, was determined. This process began by formulating standardized data collection cycle and source categories. Based on when data for each sub-parameter should be collected, the following data collection cycle categories were determined: daily, weekly, monthly, initially and with crew change, and initially. Each sub-parameter was assigned a data collection cycle based on its particular nature, which was established using expert opinions collected using factor surveys (discussed in the following chapter), which evaluated not only the impact but also the existence/frequency of the parameters influencing CLP, together with assumed data variability. Of the 314 detailed sub-parameters, 36 were assigned to be collected daily, 63 weekly, 52 monthly, 7 initially with a change in crew, and 156 initially. The detail of the data collection cycle of the sub-parameters at the different hierarchy levels is shown in Fig. 2.3.

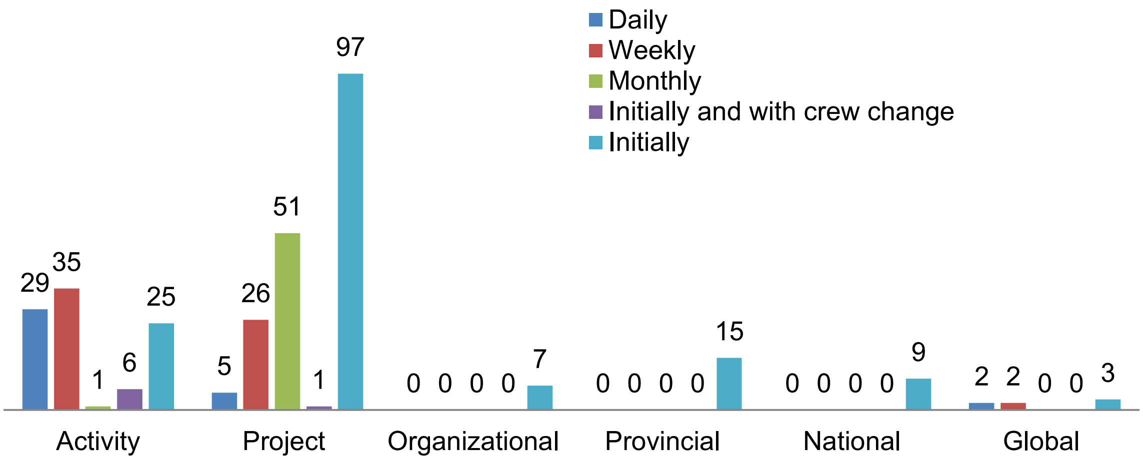

Figure 2.3: Input Sub-parameters Data Collection Cycle

The following data source categories were determined: crew members, foreman, superintendent, project manager, and researcher or data collector. Each sub-parameter was assigned a data source based on which respondent type would have the best knowledge of the state of the sub-parameter. Using multileveled factor surveys and expert opinions, each sub-parameter's level within the project hierarchy was verified, and sub-parameter's level was used to determine its data source in terms of the most appropriate project member to be targeted so as to gather accurate values; detail discussion of the process is presented in the following chapter. Generally, for sub-parameters defined at the activity level, crew members and the foreman were identified as the most appropriate sources to target for data collection, while for sub-parameters defined at the project level and higher, the superintendent and project manager were identified as the best sources. Sub-parameters whose values can be collected from direct observation or other accessible sources, like weather databases, were assigned to the researcher or data collector. The detail of the data source categories of the sub-parameters together with the data collection cycles is shown in Fig. 2.4. The detail results of the input sub-parameters quantification process is discussed in the following sub-sections based on the hierarchal levels of the parameters and their respective sub-parameters.

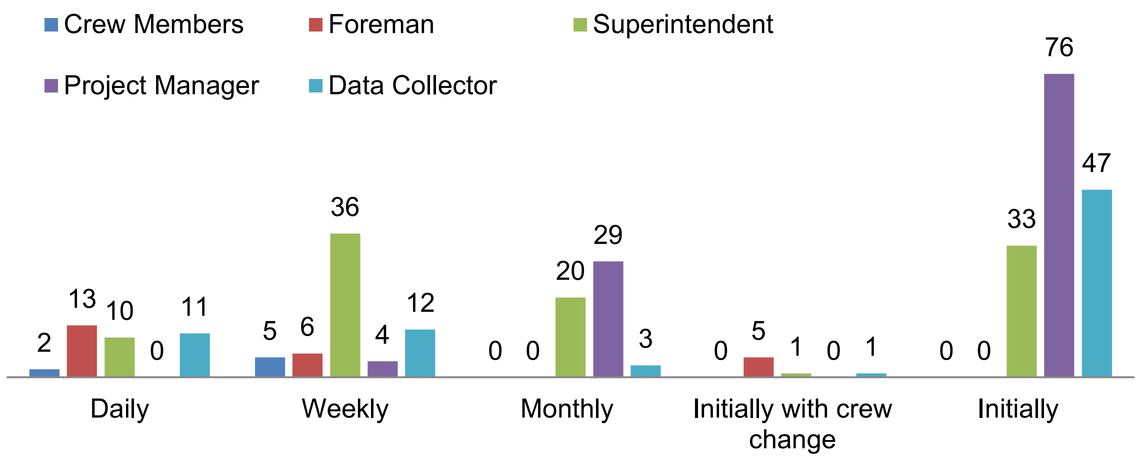

Figure 2.4: Input Sub-parameters Data Sources and Data Collection Cycle

2.3.1.1: Activity Level Input Parameters

A total of 96 sub-parameters were developed characterizing the identified activity level input parameters shown in Table 2.2. As shown in Fig. 2.5, the activity level input parameters were further grouped under the following categories:

- 1.1. Labour characteristics,
- 1.2. Material,
- 1.3. Equipment and tool
- 1.4. Task property,
- 1.5. Task location property,
- 1.6. Foreman, and
- 1.7. Engineering and instructions.

The full details of the developed input sub-parameters together with description, scale of measure, data collection cycle, and data source are shown in Appendix A.2. Selected examples of activity level parameters are shown in Table 2.3. The data collection cycles have been abbreviated as daily (D), weekly (W), monthly (M), initially and with crew change (C), and initially (I). The data source have been abbreviated as crew members (CM), foreman (FM), superintendent (SI), project manager (PM), and data collector (DC).

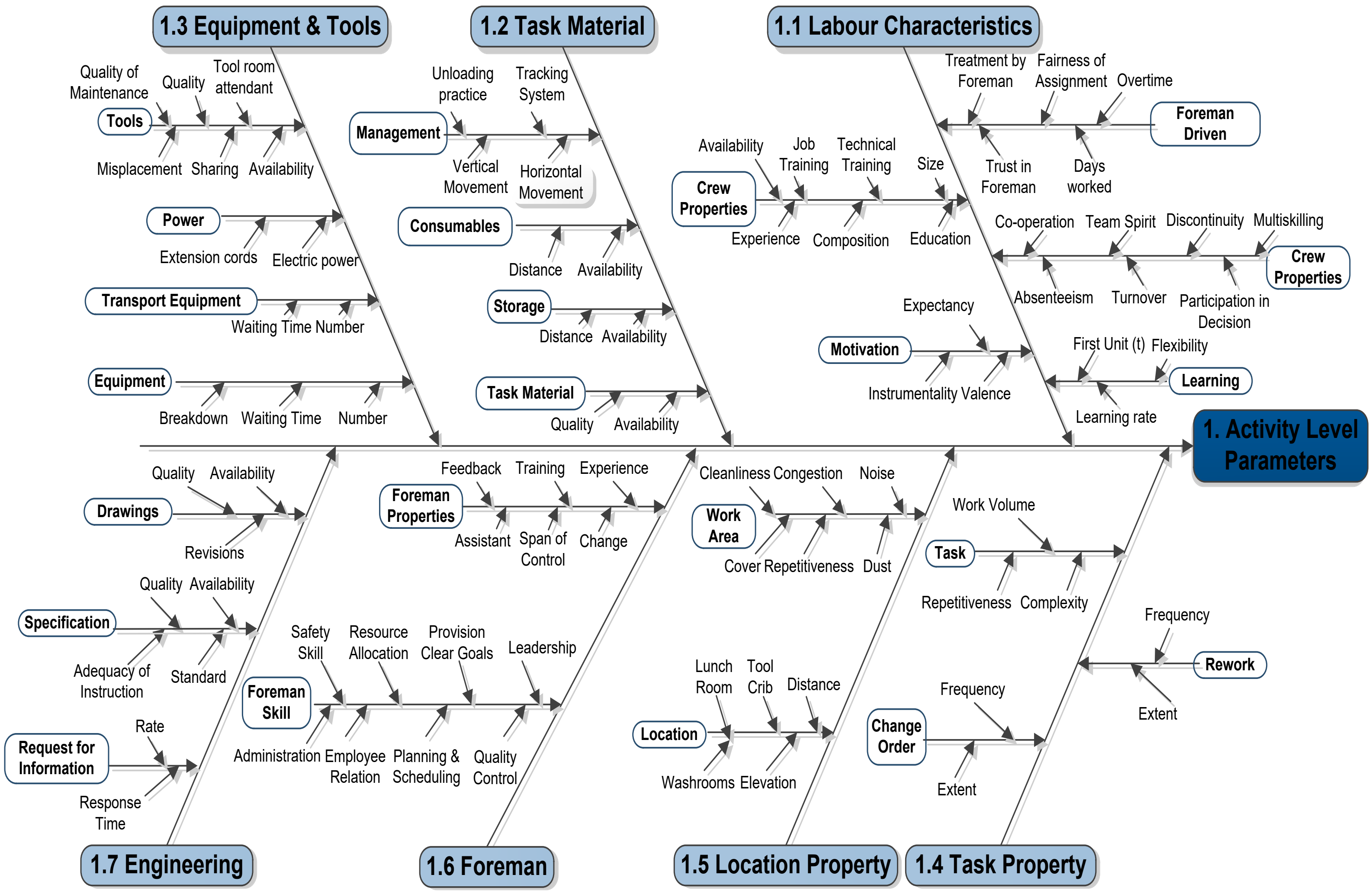

Figure 2.5: Activity Level Input Parameters and Sub-parameters

Table 2.3: Activity Level Input Parameters Quantification

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.1.1	Crew Properties	Refers to the nature and property of the crew and its members which will directly involve with execution of the tasks.		D	FM
1.1.1.1	Crew size	The total size of the crew performing the actual task will have a direct effect on the amount of output.	Integer (Total number of crew members)	D	DC
1.1.1.2	Craftsperson education	Refers to the highest achieved education level of craftsperson in a crew. The most common education level of the crew members is recorded.	Categorical (Below Secondary, Secondary School, Technical or Apprentice, College, University)	C	DC
1.1.1.3	Craftsperson on job training	Craftspeople are expected to get job specific trainings to improve their skillset. Any training, for erecting scaffolding, rigging and hoisting, zoom boom operation, etc. provided to craftspeople during his/her career is recorded. The average training hour per crew is recorded.	Real number (No. trainings attended x Duration of Training, hrs.)	C	FM
1.1.12	Fairness of work assignment	Refers to the feeling of the crew members towards the assignment of work by foreman to the different crews and crewmembers. It will be measured in terms of consistency (same policy), reasonableness (use of common sense), and information (provision of information)[a]	1 - 5 Predetermined rating (shown below)	D	CM
		1. Inconsistent work assignment on a daily basis, Unreasonable work assignment among crew members, VERY POOR Information provision; 2. Inconsistent work assignment on a daily basis, Unreasonable work assignment among crew members, POOR Information provision; 3. SOMEWHAT Consistent work assignment on a daily basis, Reasonable work assignment among crew members, AVERAGE Information provision; 4. VERY Consistent work assignment on a daily basis, Reasonable work assignment among crew members, GOOD Information provision; 5. VERY Consistent work assignment on a daily basis, Reasonable work assignment among crew members, VERY GOOD Information provision			
1.2.10	Material movement practices (horizontal)	Refers to the horizontal distance between the site main storage and the location where the task is being executed.	Real Number (average distance, m)	W	DC

Source: [a] Sheppard, B. H., and Lewicki, R. J. (1987). “Toward general principles of managerial fairness.” *Social Justice Res.*, *1*(2), 161-176.

2.3.1.2: Project Level Input Parameters

A total of 180 sub-parameters were developed characterizing the identified project level input parameters shown in Table 2.2. Selected examples of project level parameters are shown in Table 2.4. The details of the developed sub-parameters together with description, scale of measure, data collection cycle, and data source are shown in Appendix A.3, where the project level input parameters were further grouped under the following categories, as shown in Fig. 2.6:

- 2.1. Project delivery, contract, design documents,
- 2.2. Project nature,
- 2.3. Project condition,
- 2.4. Project owner,
- 2.5. Project team,
- 2.6. Project labour and union,
- 2.7. Salary and benefits,
- 2.8. Project initial planning practice,
- 2.9. Project scope management practice,
- 2.10. Project time management practice,
- 2.11. Project cost management practice,
- 2.12. Project quality management practice,
- 2.13. Project procurement management practice,
- 2.14. Project safety management practice,
- 2.15. Project risk management practice,
- 2.16. Project communication management practice,
- 2.17. Project human resource management practice,
- 2.18. Project environmental management practice, and
- 2.19. Project claim management practice categories.

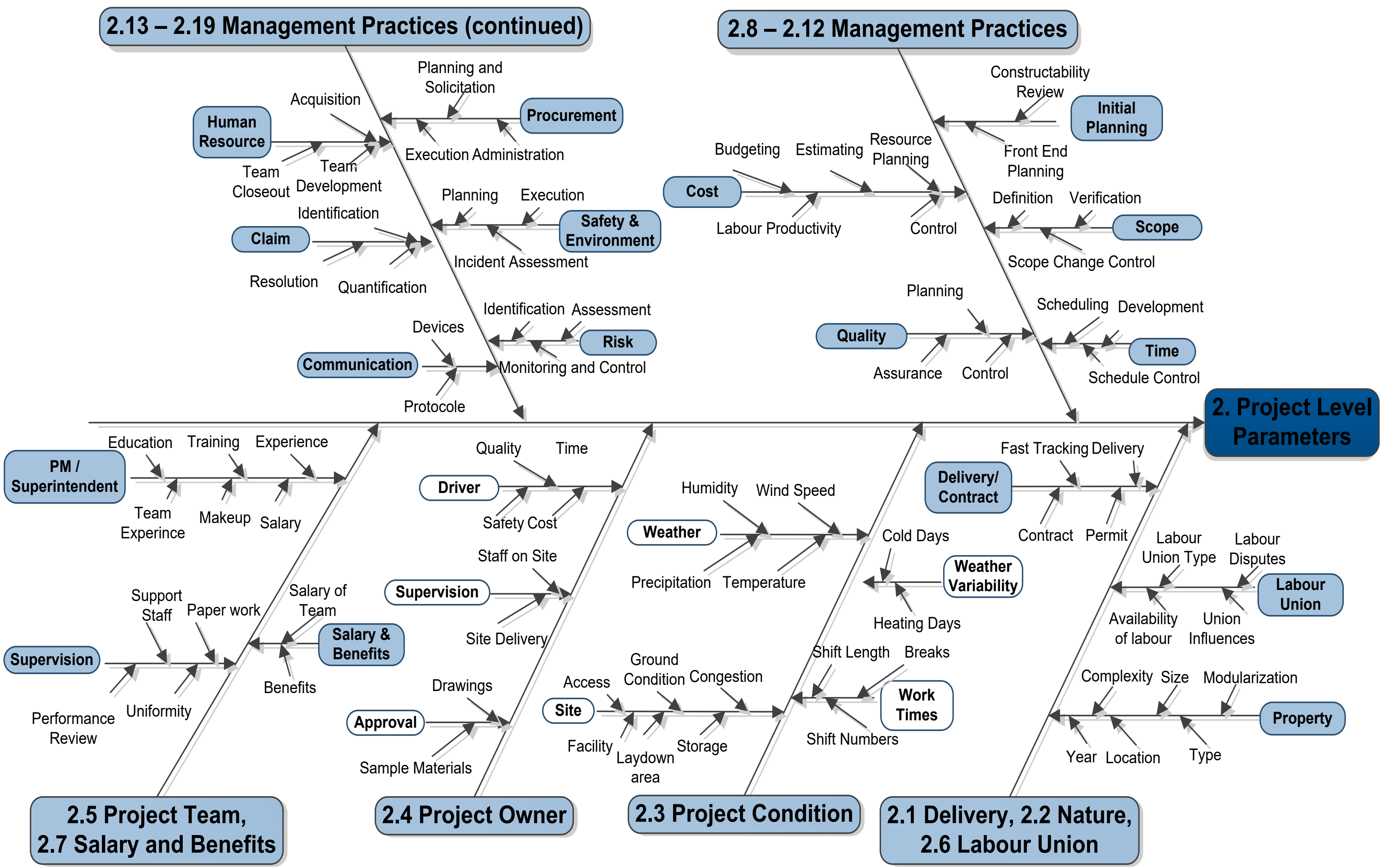

Figure 2.6: Project Level Input Parameters and Sub-parameters

Table 2.4: Project Level Input Parameters Quantification

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.1.1	Project delivery system	Refers to the arrangement between the owner and contractor on the means to design, execute, and operate the project.	Categorical (Design bid build, Design build, Build operate transfer, Private public partnership)	I	SI
2.2.1	Project type	Self-explanatory.	Categorical (Commercial, Institutional, Residential, Industrial)	I	DC
2.1.3	Level of fast tracking	Refers to whether the project construction begun before the completion of the design process.	Real number (% Overlap between design and construction schedule)	I	PM
2.2.3.3	Construction methods and techniques	Refers to the construction methods and techniques adopted for the project and the experience and availability of proper procedure with methods and technologies.	1 - 5 Predetermined rating (shown below)	I	PM
		1. VERY POOR Experience with methods and technology, LACK of proper procedure; 2. POOR Experience with methods and technology, LACK of proper procedure; 3. FAIR Experience with methods and technology, WITH proper procedure; 4. GOOD Experience with methods and technology, WITH proper procedure; 5. VERY GOOD Experience with methods and technology , WITH proper procedure			
2.3.3	Weather (temperature)	Refers to the recorded temperature at 1:00 PM of the work day.	Real number (˚C)	D	DC
2.4.1	Owner staff on site	Refers to the total number of owner staff on site to supervise the project works.	Integer (Total number of owner staff on site)	I	PM
2.5.11	Project Manager experience	Self-explanatory.	Real number (years of experience)	I	PM
2.5.6	Uniformity of work rules by superintendent	Self-explanatory.	1 - 5 Predetermined rating (shown below)	W	PM
		1. VERY Irregular among crews and HIGHLY Variable in daily work times and work days; 2. Irregular among crews and Variable in daily work times and work days; 3. Uniform among crews and Variable in daily work times and work days; 4. Uniform among crews, Always the same in daily work times and work days; 5. VERY Uniform among crews, Always the same in daily work times and work days			
2.11.2.2	Estimation team experience	Self-explanatory.	Real number (Average years of experience of estimation team)	I	PM
2.11.2.3	Time allowed for estimation	Self-explanatory.	Integer (Time taken for estimation, working days)	I	PM

2.3.1.3: Organization Level Input Parameters

A total of seven sub-parameters were developed characterizing the identified organization level input parameters shown in Table 2.2. The details of the developed organizational level sub-parameters together with description, scale of measure, data collection cycle, and data source are shown in Appendix A.4. Selected examples of organization level parameters are shown in Table 2.5.

2.3.1.4: Provincial Level Input Parameters

A total of 15 sub-parameters were developed characterizing the identified provincial level input parameters shown in Table 2.2. The details of the developed provincial level sub-parameters together with description, scale of measure, data collection cycle, and data source are shown in Appendix 4.5. Selected examples of provincial level parameters are shown in Table 2.5.

2.3.1.5: National Level Input Parameters

A total of nine sub-parameters were developed characterizing the identified national level input parameters shown in Table 2.2. The details of the developed national level sub-parameters together with description, scale of measure, data collection cycle, and data source are shown in Appendix 4.6. Selected examples of national level parameters are shown in Table 2.5.

2.3.1.6: Global Level Input Parameters

A total of seven sub-parameters were developed characterizing the identified global level input parameters shown in Table 2.2. The details of the developed global level sub-parameters together with description, scale of measure, data collection cycle, and data source are shown in Appendix 4.7. Selected examples of global level parameters are shown in Table 2.5.

Table 2.5: Organization, Provincial, National, and Global Level Input Parameters Quantification

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
3.1	Organization's principal project type	Defines the types of industries or projects types the organization is seeking	Categorical (Industrial, Commercial, Infrastructure, Institutional, Other)	I	PM
3.2	Organization year in industry	This factor indicates the number of years an organization has been in operation.	Real number (Years in industry)	I	PM
4.2	Total number of project within province	Refers to the number projects similar to the project under study which will compete for resource in the province.	Integer (Number of projects under construction per year in province)	I	DC
4.3	Provincial codes and regulations	Refers to the flexibility of provincial codes and regulations towards the construction industry.	1 - 5 Predetermined rating	I	DC
		1. Most restricted regulations, 2. Strict regulations, 3. Normal regulations, 4. Flexible regulations, 5. Most flexible regulations			
5.1	Political system	Refers to relative stability of the Canadian political system during the initiation and execution of the project.	1 - 5 Predetermined rating	I	PM
		1. VERY unstable; 2. Unstable; 3. Stable; 4. Stable; 5. Very Stable			
5.2	Foreign workers recruitment	Refers to the execution of the foreign workers recruitment program in terms of strictness and processing times.	1 - 5 Predetermined rating	I	DC
		1. VERY STRICT regulations, VERY LONG process time; 2. STRICT regulations, LONG process time; 3. NORMAL regulations, FAIR process time; 4. Flexible regulations, SHORT process time; 5. VERY flexible regulations, VERY SHORT process time			
6.1	Global economic outlook	Refers to the national economic outlook in terms of real GDP for the coming year, based on the IMF world economy outlook [b]	Real number (Real GDP growth, %)	I	DC
6.3.1	Oil price	The average WTC (Western Texas Intermediate) oil price is recorded.	Real number (Dollar / barrel)	D	DC
6.3.2	Price fluctuation	The average net price fluctuation on a weekly basis is recorded.	Real number (Weekly price change, %)	W	DC

Source: [b] IMF (2015). “World economic outlook.” International Monetary Fund, < http://www.imf.org/external/pubs/ft/weo/2015/01/ > (June 28, 2015).

2.3.2: Identification and Quantification of Process Parameters Influencing CLP

In the construction industry, understanding how time is used during the input-to-output conversion process is also vital to understanding CLP; work-study methods are commonly employed for this purpose. Existing work-study data collection methods related to CLP include: work sampling (activity sampling), field rating (five-minute rating), process analysis, time studies, predetermined motion time systems, method productivity delay model, and crew balance charts (Drewin 1982; Dozzi and AbouRizk 1993). Dozzi and AbouRizk (1993) also suggested the use of field surveys (craftsman questionnaires and foreman delay surveys) to gather information on causes of delays, obstacles to the work, and possible improvements.

Work sampling (WS), a method used to determine the amount of time workers spend performing direct (productive), support, or delay work proportions, is the most widely used work-study method (Picard 2004; Josephson and Björkman 2013). Some advantages of WS include its suitability for activities that are non-repetitive (Liou and Borcherding 1986). WS is among the few quantitative tools that can show inefficiencies and problem areas, it is also simple, requires minimal resources to conduct, and provides quick results. As a procedure, it is less intrusive as compared to other work studies, such as time-lapse photography, and focuses on crew-level measurements (Liou and Borcherding 1986; Drewin 1982; Al–Ghamdi 1995). Its disadvantages include its inability to reveal the sources of inefficiency and differences between individuals, plus its inability to provide a reliable productivity estimate (Drewin 1982; Al–Ghamdi 1995).

In the 1970s and 1980s, work sampling gained increased application in the construction industry (Thomas et al. 1984). Craftsman questionnaires and foreman delay surveys have also been used to both supplement and complement it (Chang and Borcherding 1985; Gouett et al. 2011). Recently several studies have demonstrated its application to productivity improvements (Gouett et al. 2011). However, since WS studies do not measure actual output, such as units installed, it does not assess actual labour productivity (Thomas et al. 1990). The use of matching productivity data with a crew-level WS study can improve on its shortcomings to clearly indicate the actual efficiency of the crew. The usefulness of such studies, however, has yet to be reliably established (CII 2010). In part, this is because past WS studies have focused on decreasing the delay (ineffective) proportion of activities, in order to maximize direct work (tool time), based on the assumption that an increase in direct work will increase productivity (Thomas 1991; Tsehayae and Fayek 2012).

Work sampling proportions summarize the actual utilization of labour work hours using proportions of work time spent on performing work categories like direct work, support, etc., and provide an in-depth examination of what happens during the conversion process. Past studies have shown that the definition of WS categories and the subsequent task classification can significantly affect the accuracy of the WS proportions, and, hence, their relationships with CLP (Thomas 1991). However, past work sampling studies had different and mostly not clear aims, resulting in varied definition of WS categories (Josephson and Björkman 2013). Hada and Abdalla (1989) employed 10 WS categories—direct work, receiving instructions, tools and materials, transportation, waiting, travel, breaks, personal, late starts, and unexplained, to study a house framing activity. Josephson and Björkman (2013) adopted three lean focused WS categories—direct value-adding work (representing tasks physically adding to the product); preparations (representing tasks which cannot be immediately eliminated without affecting the customer value, but still do not add value); and waste (representing tasks if eliminated would not affect the customer value).

In WS studies, observers make continuous judgments in identifying the tasks a worker is carrying out and assign them to any one of the work sampling categories. Thus, it is imperative that the WS categories are kept to the minimum and the definitions of the WS categories are standardized (Josephson and Björkman 2013). CII (2010) presented the *Guide to activity analysis,* which formalized and standardized the procedure for conducting work sampling studies, and developed standard seven WS categories—direct work, preparatory work, tools and equipment, material handling, waiting, travel, and personal to study any construction activity (refer to Table 2.6). Thus, these seven WS categories are adopted in this study as process parameters (*P*), and are used to properly examine what happens during the conversion process, and develop the input-process-output CLP models. In Table 2.6, the identification labels of each process parameter are also included.

Table 2.6: Process Parameters (Work Sampling Proportions) Influencing CLP

Source	WS Categories	Tasks
CII 2010	Direct Work $_{y1}$	Exertion of physical effort directed towards an activity or towards assisting in an activity; involves workers installing materials, but also includes the physical effort of support groups
	Preparatory Work $_{y2}$	Activities related to receiving assignments and determining requirements prior to performing tasks. Preparatory work includes stretching activities, safety talks, start-card process, and discussions to explain or plan the assigned task at the work location. These discussions can take place between craft workers or between supervisors and craft workers.
	Tools and Equipment $_{y3}$	Activities associated with obtaining, transporting, and adjusting tools or equipment in preparation of performing direct work
	Material Handling $_{y4}$	Includes the transportation of materials from one part of the facility to another, not including items moved in the general area of the task or into their final position
	Waiting $_{y5}$	Periods of waiting or idleness, even if workers are attentive to on-going work by others
	Travel $_{y6}$	Walking or riding either empty-handed or without tools, materials, or technical information
	Personal $_{y7}$	Time taken or periods of idleness during normal work-hours; workers taking personal time are normally not attentive. This category excludes normal unpaid breaks like lunch periods but includes paid breaks like coffee breaks.

2.3.2.1: Quantification of Process Parameters

In this study, the process parameters were quantified using the following three steps: defining the WS categories, determining the observation method, and defining the sample size. The various activities under this study were represented using the work time proportions based on seven work categories; for each WS category, the corresponding Construction Industry Institute (CII 2010) category names and definitions shown in Table 2.6 were applied. The proportion of time spent on each WS category is calculated for each crew observed for a certain period of time and based the total head counts for each WS category. For example, the direct work-time proportion, also known as “tool time” or “wrench time”, represents the amount of time spent actively producing units of output that contribute to the completion of the project components. The direct work-time was calculated as shown in Eq. (2.2). Similarly, the work proportions for each of the six other WS categories were calculated by taking the ratio of their respective number of observations to the total number of observations.

$$\text{Direct Work Proportion} = \frac{\text{Total Observations of Direct Work}}{\text{Total Number of Observations}} \qquad (2.2)$$

In WS studies, three observation methods, namely the tour, crew, and modified crew methods are recommended (CII 2010). Accordingly, the tour observation method studies all craft so as to establish the overall process efficiency of the entire site; while the crew observation method focuses on an individual crew. The modified crew observation method extends the crew approach by studying the entire craft and sampling representative crews performing the craft, resulting in improved results over the crew method (CII 2010). Thus, taking electrical craft as example, a subset of all electrical crews, performing activities like wire pulling, switch installation, piping, etc. will be studied (CII 2010). In this study, the modified crew observation method is adopted, and the WS observation focuses on the study of representative crews performing selected activities.

In order to achieve statistically significant process parameters, a sufficient number of random observations must be made for each of the crews performing the activities under study. Traditionally a binomial distribution sample size, shown in Eq. (2.3), is used to determine WS sample size n (Dozzi and AbouRizk 1993; Aft 2000), where $Z_{\alpha/2}$ represents the standard normal variable corresponding to a given confidence level and p represents the category percentage:

$$\mathrm{n} = \frac{(Z_{\alpha/2})^2\, p\,(1-p)}{d^2} \qquad (2.3)$$

In WS studies the use an error of ±5% at a confidence level of 95% is recommend (CII 2010), resulting $Z_{\alpha/2} = 1.96$. Considering the worst case category percentage ($p = 50\%$), the total binomial WS sample size would be equal to 384 (Dozzi and AbouRizk 1993). However, as there are seven WS categories in study, a multinomial distribution sample size, shown in Eq. (2.4), was used to determine sample size n (Gouett et al. 2011). The equation was calculated at varying numbers of categories

m (regardless of the number of categories in particular WS study) to find the maximum number of observation n in the worst case scenario.

$$\mathrm{n} = \max\left[\frac{(Z_{(1-\frac{\alpha}{2m})})^2 \; \frac{1}{m} \; (1-\frac{1}{m})}{d^2}\right] \tag{2.4}$$

Thus, using Eq. (2.4) and for a 95% confidence level and error of $d = 0.05$, the maximum number of observations n in the work case scenario results in $m = 3$, and $n = 510$ observations per hour of the study. Additionally, it is desirable to study WS category proportions for each 1-h period of a workday, so for 8-h workday, there will be 8 study periods and resulting in a total of 4,080 observations. However, not all 510 observations have to be made during a single hour, and the study could be completed over a certain period of time, thereby reducing the required number of observation (CII 2010). Therefore, the length of WS study depends on the sample size as only finite number of observations can be collected in a day, and it is recommend that such study shall last at least one to three weeks (CII 2010). In this study, WS observations were complemented by foreman delay surveys and craftsman questionnaires, which are used to quantify the causes of delays for each crew under study.

2.3.2.2: Foreman Delay Surveys

The purpose of the foreman delay surveys is to record the causes of delays for each crew under study, as a foreman is the person most familiar with the crew and the problems that cause delays (Dozzi and AbouRizk 1993). Use of foreman delay surveys have shown that reasonably accurate information can be collected on causes of delay, and aid the measurement and improvement of productivity (Oglesby et al. 1989). Foreman delay surveys are collected either daily for several weeks, daily with weekly reporting, or as needed (CII 2013). In this study, such surveys are collected daily, and at the end of the daily data collection process, the foreman estimated the number of hours lost due to the delay sources included in the survey, for each of the crews under observation. Each lost hour value is multiplied by the number of workers affected and then summed to determine the overall lost work-hours of each shift.

2.3.2.3: Craftsman Questionnaires

The purpose of the craftsman questionnaire is to identify the occurrence of common factors that inhibit the productivity of the craft workers and to estimate the work-hours lost per craftsman per day (Dozzi and AbouRizk 1993; Hanna 2010). Craftsman questionnaires are useful for identifying the possible reasons for inefficiencies, delays, and available solutions, and offer a means of verifying other management reporting procedures and their results (Chang 1986; Rivas et al. 2011; CII 2013). Craftsman questionnaires are collected either daily for several weeks or when needed (CII 2013). In this study, the questionnaire is completed daily in an interview format with a randomly selected crew member.

The lost hours values determined using the foreman delay surveys were deducted from the crew total labour work-hours, see Eq. (2.3), used to calculate CLP, if verified by crew members via the craftsman questionnaires; the lost hours values also provide an explanation for WS results associated with delay events. If, for example, a crew worked for 8 hours in a shift and had a delay due to weather amounting to 1 hour (verified in both the foreman delay survey and craftsman questionnaire collected at the end of the shift), the work hours would be reduced to 7 hours and the total work-hours used to calculate productivity would be computed as a product of 7 hours and the crew size.

2.3.3: Quantification of Output Parameter or CLP

The efficiency of activity level systems, focusing on the labour crew of the construction process, is measured using construction labour productivity (CLP). In this thesis, CLP is defined as the ratio of units of output to units of input work hours—as shown in Eq. (2.5), where higher values are better than lower values.

$$\text{Construction labour productivity (CLP)} = \frac{\text{Output (installed quantity)}}{\text{Total labour work-hours}} \qquad (2.5)$$

According to CII (2013), for crew-output variable, the units of measure for reporting quantities should be simple, easy to apply, and accurate. Several quantity measurement methods such as, units completed, percent complete, level of effort, incremental milestone, and start/finish percentage are used to quantify the output, in terms of installed quantities. Physical measurement of units completed is the preferred output measurement, and is applicable for activities with well-defined scope, few subtasks, single craft or trade, short duration for completing each unit of output, and installed quantity can be quickly determined by counting or using elementary math (CII 1990); and as in this study, the activities under investigation had met these requirements, and physical measurement of units was used to determine the installed quantities. The total labour work-hours were recorded based on the total man-hours the crew worked including scheduled breaks but minus lunch time and the recordings were made by the data collector.

2.3.5: Data Collection Methodology

Construction labour productivity is a cause of great concern for both the construction industry and academia, however CLP measurement and modeling is yet to be fully standardized. Construction companies mainly focus on cumulative average productivity value and ensuring the estimated level of productivity is achieved or bettered, while researchers focus on “average level of productivity during much shorter periods of time when a particular set of conditions exist” (Thomas et al. 1990). This research aims to develop interpretable and accurate context-specific and universal CLP models based on the *Input-Process-Output* system variables. Thus, in order to achieve this goal, labour intensive project activities related to four trades: concreting, electrical, mechanical, and boilermakers were studied. The four trades are among the most labour intensive trades in the construction industry, and have been among the most studied trades in past CLP researches (Koehn and Brown 1985; Sonmez and Rowings 1998; Hanna et al. 2005; Ezeldin and Sharara 2006; Silva and Ruwanpura 2006; Nasirzadeh and Nojedehi 2013; Heravi and Eslamdoost 2015). Selected activities for each trade were studied with a number of construction companies, with at least three projects per company. Thus, projects under different contexts, based on project types, location, etc. were examined, so as to properly capture the effect of context on CLP.

This research data collection process had four main steps, which involve: (1) identifying candidate projects from construction companies; (2) conducting site kick-off meeting to introduce the study to the research to participants; (3) conducting factor survey to identify critical input parameters (factors and practices); and (4) field data collection on system parameters made up of labour productivity, work sampling, and input parameters. A data collection methodology having two main data collection components was developed to carry out the study of the activities. In the first step, the research began by identifying three candidate projects from a number of construction companies, through a research recruitment meeting. The three projects were required to have activities for the principal trades under study: concreting, electrical, mechanical, or boilermakers. The criteria for identifying projects, activities, and crews for this study are summarized as shown in Table 2.7.

Table 2.7: Project, Activity, and Crew Selection Criteria

Item	Criteria	Preference
Project	- Include the principal trade types for study, and has activities that extend for at least three months - Shall be within a close proximity to city of Edmonton	Three projects per company
Activity	- Activities that are labour intensive and have higher man-hours for the trade under study - Quantifiable outputs in terms of installed quantities - Common among several projects - Executed by at least three crews	Preferably carried out by multiple crews
Crew	- Crews for study shall be randomly selected from the existing crews performing similar activities	Three crews

The activities were also required to extend for a minimum of three data collection cycles, and were executed by at least three crews. The crews performing the activities were randomly selected, based on similarity of assigned activities. A data collection cycle was considered to be of a month period, and when needed, was extended to provide additional data sets. The three data collection cycles were positioned in such a way that the activity was examined under different weather seasons. Once the projects and activities were identified, general information on the construction company and the candidate projects was gathered. In the second step, the research commenced through a site kick-off meeting for each of the identified projects in the presence of the researcher, project manager, supervisors, foreman, and crew members. The kick-off meeting was intended to provide research participants the background of the research, the data collection process, and most importantly the research ethics procedure. The third step involved conducting factor surveys for the candidate projects, so as to identify the critical factors and practices affecting CLP based on the opinion of project respondents. The respondents were identified from project management (construction manager, project manager, project coordinator, superintendent, and supervisor) and trade (foreman and craftsperson) levels. Following the factor surveys, the last step involved the field data collection for the trades and activities under study.

2.3.5.1: Data Collection Protocol

To facilitate and standardize the field data collection of the three system variables, a comprehensive data collection protocol was developed (Tsehayae and Fayek 2014b). The protocol describes the steps and details for the field data collection of output parameters (counting work hours and measuring quantities installed), process parameters (carrying out work sampling together with delay surveys), and input parameters (documenting the state or presence of various factors and practices that can potentially affect labour productivity). Data collectors were required to read and understand the protocol, as the quality and consistency of data was of a paramount importance. The data collection protocol was based on 10 data collection forms, as shown in Table 2.8, and included detail presentation of the forms and instructions on how the forms should be filled.

Table 2.8: Summary of Data Collection Forms

Form	Description	Data Collection Frequency
1	Productivity Data	Daily
2	Work Sampling	Daily
3	Foreman Delay Survey	Daily
4	Craftsman Questionnaire	Daily
5	Daily Input Variables	Daily
6	Weekly Input Variables	Weekly
7	Project Diary	Daily
8	Monthly Input Variables	Monthly
9	Project Features	Initially and when there is a change in crew members
10	Context Variables	Initially

2.3.5.2: Productivity Tracker Database

To facilitate the comprehensive data collection and analysis required for modeling CLP, a database application tool, ProductivityTracker©, was developed. The tool has a setup, data inventory, and report/analysis modules (see Fig. 2.7), and stores the CLP system variables data from all construction projects studied. Additionally, a help module was developed to facilitate data entry and report generation. The tool's security setting allowed users different levels of access in order to preserve the data anonymity and security.

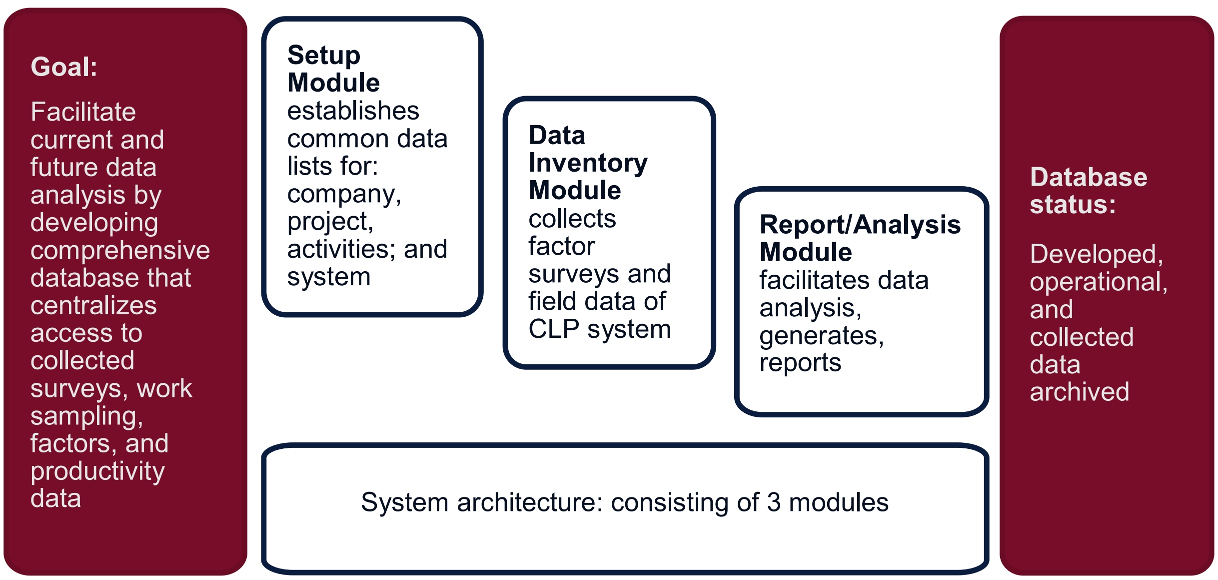

Figure 2.7: ProductivityTracker: Database Architecture

The setup module consists of three main components: a setup module for adding the details of companies, projects, activities, and the crews under study (see Fig. 2.8); a setup module for defining the input parameters for factor surveys and field data collection (see Fig. 2.9); and a security module for creating user lists, and also defining roles of users as administrator, data analyzer, or data entry. The data inventory module consists of factor survey module (see Fig. 2.10), and field data module (see Fig. 2.11), used for adding the collected factor surveys and field data on system variables of CLP. The report module generates reports from the collected data.

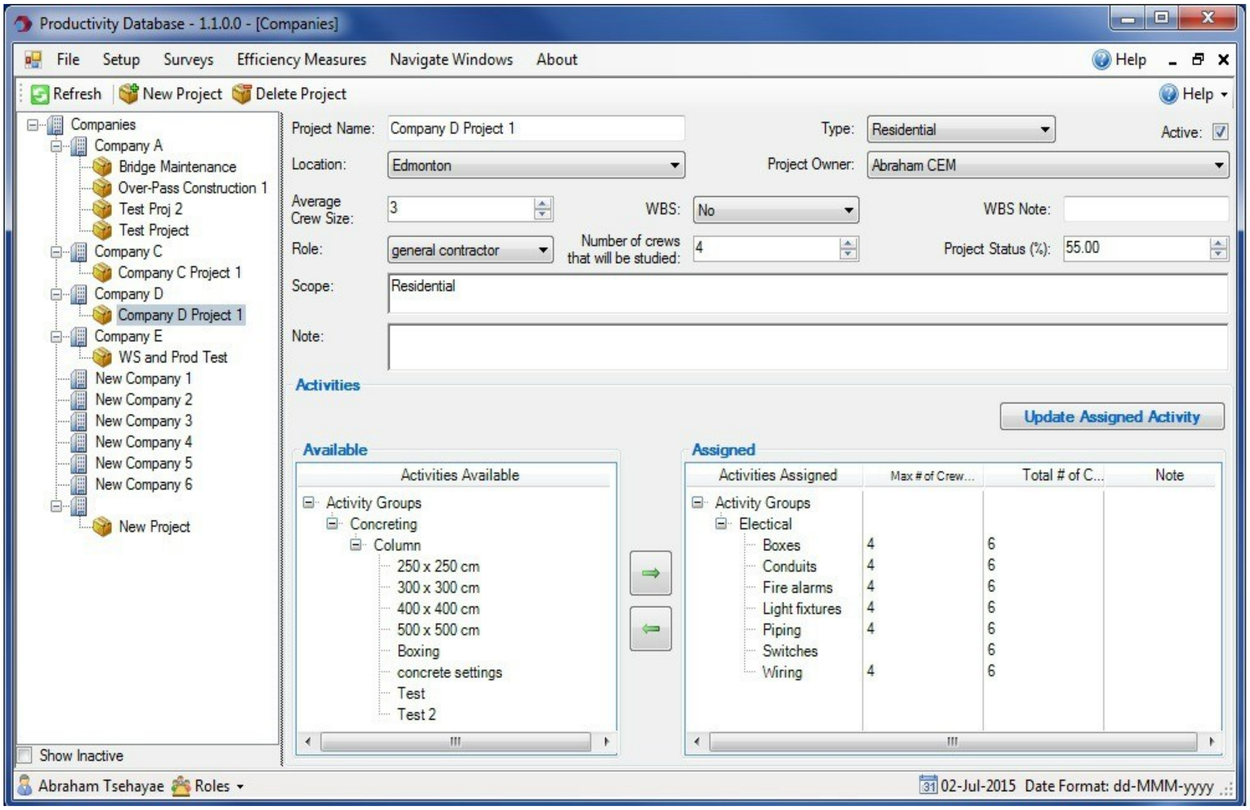

Figure 2.8: ProductivityTracker: Setup for Companies, Projects, and Activities

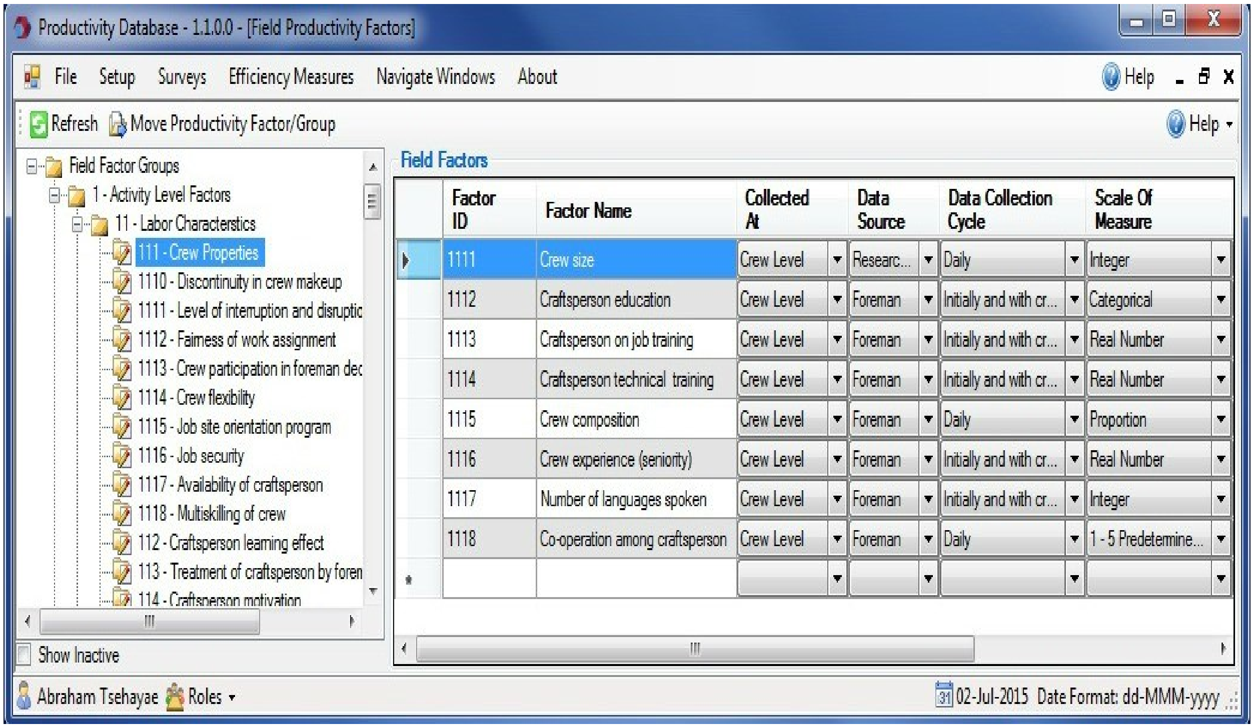

Figure 2.9: ProductivityTracker: Setup for Input Parameters (Factors and Practices)

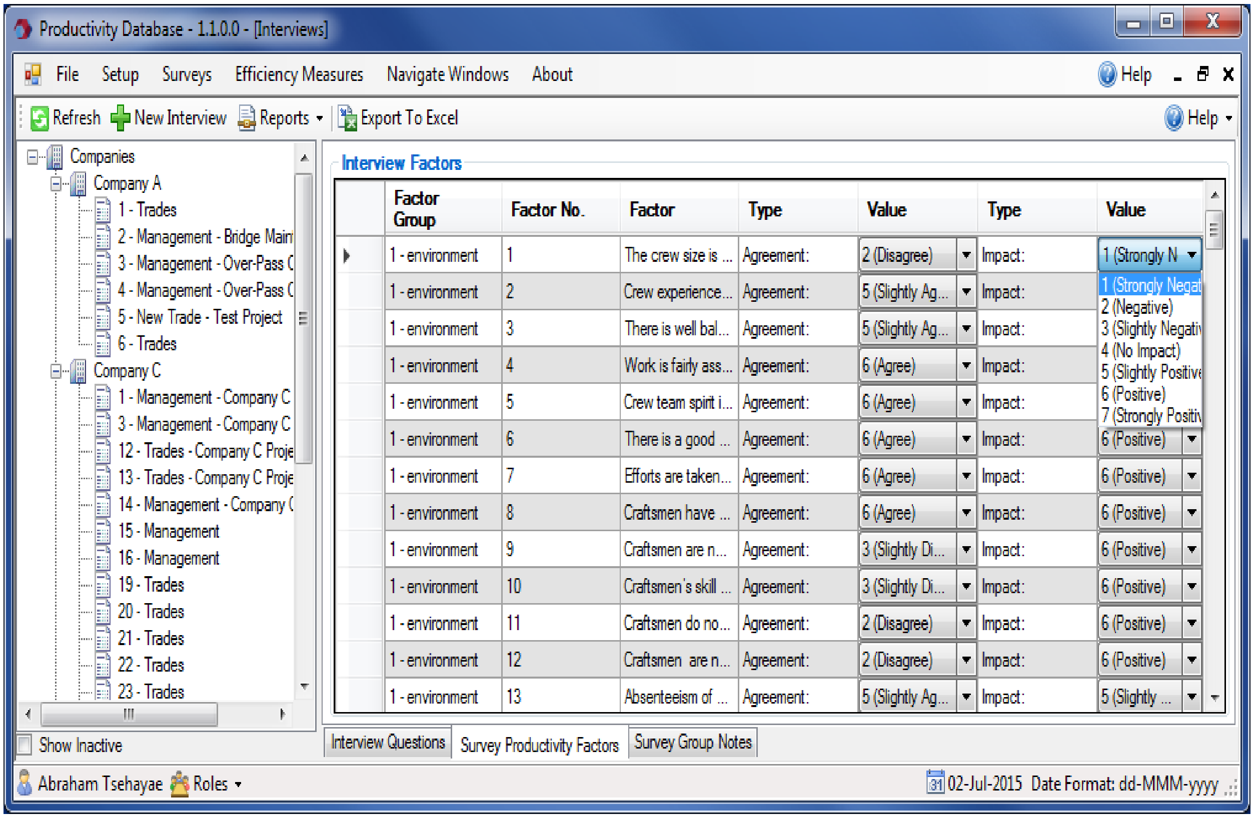

Figure 2.10: ProductivityTracker: Factor Survey Module

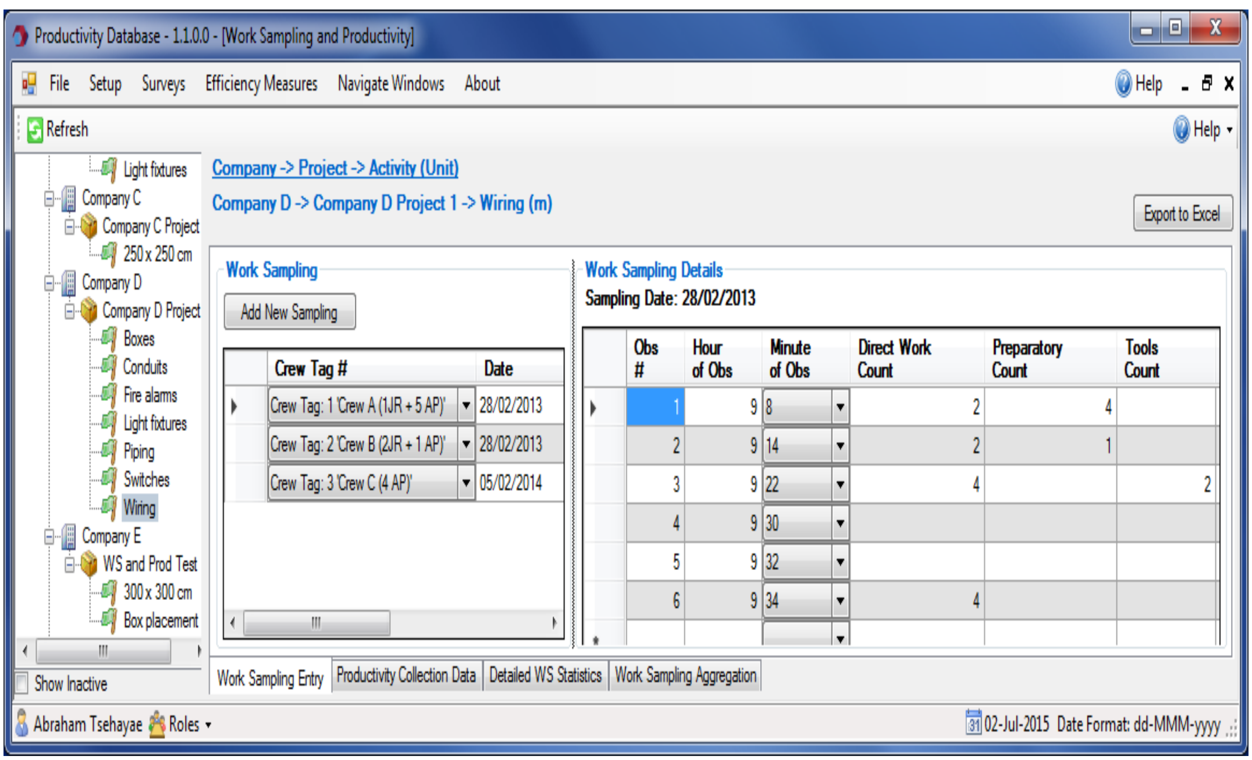

Figure 2.11: ProductivityTracker: Field Data Module

2.3.5.3: Research Ethics Procedure

The study presented in this thesis involves the study of construction crews converting inputs to outputs, and implicates living human participants or their data. Accordingly, a research ethics procedure was developed and approved by the Research Ethics Office of University of Alberta. The research information collected from the several construction projects was kept strictly confidential using data encryption and locked filling cabinets. Appropriate safeguards were put in place for data collection, use, and dissemination. Additionally, research results were released if and only if there is no risk of participant (individuals or companies) re-identification.

For the factor survey data collection stage, the following research ethics procedure was adopted: participants were informed of the study goals and the consent process (information letter and consent form), and written consent was documented; also, appropriate environment was selected in conducting the interview survey so as to protect the privacy of the participant, and participants were given the right to withdraw (2 weeks from date of survey data collection). Collected surveys were then anonymized using code sheets.

For the field data collection stage, several procedures were adopted. In conducting field data collection, an anonymous random observation of crews was adopted; participants were informed of the work sampling observation during the kick off meeting, and consent was not documented as construction projects have limited expectation of privacy. However, in conducting field data collection of foreman delay surveys and craftsman questionnaires, consent was documented.

2. 4: DATA COLLECTION RESULTS

In order to gather adequate data for detailed analysis and modeling of CLP, extensive data for input, process, and output parameters were collected from 11 projects across Alberta, Canada over a 29-month period. The activities were studied in three data collection cycles, where each cycle extended over a month-long period and encompassed different weather seasons. The following trades were studied: concreting, electrical, mechanical, and boilermakers. However, limited data were available for the mechanical trade and only the factor survey data was found useful. The trade categories (total number of studied projects shown in parentheses), project type, activities studied, description of the activities, and the number of total data instances (*N*) collected is shown in Table 2.9. Data collection took place between June 2012 and November 2014.

Table 2.9: Profile of Activities Studied for CLP Modeling

Trade category	**Project types**	**Activity**	**Activity description**	***N***
Concreting (6)	Commercial mixed-use office-staff facility building, industrial warehouse building, commercial warehouse building, mixed residential-community center building, high-rise mixed commercial-residential building, institutional building	Columns	Concrete placement for columns	21
		Footings	Concrete placement for footings	5
		Grade beams	Concrete placement for grade beams	6
		Pile caps	Concrete placement for pile caps	2
		Slabs	Concrete placement for slabs	28
		Walls	Concrete placement for walls	30
Electrical (3)	Commercial mixed-use office-staff facility building, seniors residence, residential apartment	Box installation	Installation of pull & outlet boxes	48
		Conduits	Installation of flexible conduit	13
		Panel	Installation of main board panel	2
		Piping	Installation of rigid galvanized steel conduit	57
		Switches	Switch installation	6
		Wire pulling	Pulling wire	43
		Fire alarms	Fire alarm installation	5
		Light fixtures	Light fixture installation	17
Boilermaker (3)	Coal power plant boiler shutdowns	Buffing	Smoothing of tube surface before overlays	7
		Overlays	Welding additional metal layer on tubes to reinforce their thickness	38
		Shields installation	Replacing shields over tubes by removing old ones and welding new shields over tubes	68
		Tube replacement	Cutting tubes and installing and welding new ones	3

A total of 92, 191, and 116 data instances were recorded for concreting, electrical, and boilermaker trade categories, respectively. For each data collection instance, WS observations were made for the crew under study and parameters (factors and practices), total man-hours, and installed quantities were documented. For the concreting, electrical, and boilermaker activity categories, a total of 3,526; 5,108; and 6,672 random work sampling observations, respectively, were made.

The data was collected in collaboration with six partnering companies. In case of concreting activities, the first three projects were built by Company 1, a multinational construction company with over 100 years of experience, and the last three were built by Company 2, a local construction company with over 40 years of experience. In case of electrical activities, the first project was built by Company 3, a multinational electrical company with over 50 years of experience, and the last two were built by Company 4, a local electrical company with over 5 years of experience. In case of boilermaker activities, the three shutdowns were carried out by Company 6, a multinational energy company with over 70 years of experience. In case of mechanical activities, the lone project was carried out by Company 7, a local mechanical company with over 20 years of experience.

2.5: Chapter Summary

Because of its significance to project performance, CLP has been well studied. However, despite the extensive research in the area, consensus on the classification and quantification of influencing parameters, and modeling the relationship between the influencing parameters and CLP is not full achieved. This chapter first examined past CLP modeling studies, and challenges and limitations were identified. To address these limitations, this study aimed to model construction labour productivity using a system approach which involves three model parameters—*Input*, *Process*, and *Output* and developed a detailed research methodology to identify, document, and mode the system parameters under various contexts.

This chapter using critical review of CLP literature identified and classified 169 input parameters made up of influencing factors and practices. The process parameters were identified based on review of available work-study methods, and work sampling method was selected for its simplicity, minimal resource requirement, and suitability for non-repetitive activities. Accordingly, seven standard WS categories—direct work, preparatory work, tools and equipment, material handling, waiting, travel, and personal were adopted. Additionally, the process parameters were supplemented with foreman delay surveys and craftsman questionnaires to collected causes of delay. Next, an approach for the quantification of the numerous subjective and objective input and process parameters influencing CLP was carried out. For the identified input parameters, measurement scales were developed so as to quantify and enable construction site data collection. The process parameters were also quantified using the definition of the seven standard WS categories, and determining the observation method together with determination of appropriate sample size. The quantification of the output parameter or CLP was carried out using the crew-output in terms of installed quantities and total labour work-hours, and the output measurement method was presented. The details of the data collection methodology having four main steps was also presented together with the details of the custom-made, server-based database tool, called ProductivityTracker©. Additionally, the research ethics procedure developed to ensure the anonymity and confidentiality of research participants was summarized. Finally, the results of the extensive data collection process were summarized. The next chapter presents the methodology for identifying key input parameters (factors and practices) influencing labour productivity using expert and data-driven approaches.

2.6: REFERENCES

AbouRizk, S., Knowles, P., and Hermann, R. (2001). "Estimating labor production rates for industrial construction activities." *J. Constr. Eng. Manage*., 127(6), 502–551.

Aft, L.S. (2000). "Work sampling." *Work measurement and methods improvement*, 1st Ed., John Wiley and Sons, New York, US, 299–324.

Al–Ghamdi, M.A. (1995). "Evaluation of work sampling as an indicator of construction labour productivity." M.Sc. thesis, King Fahd Univ. of Petroleum and Minerals, Dhahran, Saudi Arabia.

Anand, K. B., and Ramamurthy, K. (2003). "Laboratory-based productivity study on alternative masonry systems." *J. Constr. Eng. Manage*., 129(3), 237–242.

Archiszewski, T., Skolicki, Z., and De Jong, K., 2005. "Intelligent agents: fundamentals." *Agents and multi-agent systems in construction*, Anumba, C.J., Ugwu, O.O., and Ren, Z., New York, US, 6–30.

Awad, A., and Fayek, A. Robinson. (2012). "A decision support system for contractor prequalification for surety bonding." *Autom. Constr.,* 21, 89–98.

Botta, A. (2008). "Automatic context adaptation of fuzzy systems." Ph.D. thesis, IMT Institute for Advanced Studies, Lucca, Italy.

Campbell, D.J. (1988). "Task complexity: A review and analysis." *Acad. Manage. Rev.*, 13 (1), 40–52.

Chan, A., Scott, D., and Chan, P. (2004). "Factors affecting the success of a construction project." *J. Constr. Eng. Manage*., 130(1), 153–155.

Chang, L.M. (1986). "Inferential statistics for craftsman questionnaire." *J. Constr. Eng. Manage*., 112(4), 492–499.

Chang, L.M., and Borcherding, J.D. (1985). “Evaluation of craftsman questionnaire.” *J. Constr. Eng. Manage*., 111(4), 426–437.

Chanmeka, A., Thomas, S., Caldas, C., and Mulva, S. (2012). “Assessing key factors impacting the performance and productivity of oil and gas projects in Alberta.” *Can. J. Civ. Eng.*, 39(3), 259–270.

Christian, J., and Hachey, D. (1995). "Effects of delay times on production rates in construction." *J. Constr. Eng. Manage*., 121(1), 20–26.

CII (Construction Industry Institute). (1990). *Productivity measurement: An introduction,* Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2006). *Work force view of construction labor productivity*, Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2010). *Guide to activity analysis*, Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2011). *Productivity benchmarking summary report*, Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2013). *The construction productivity handbook*, Univ. Texas at Austin, Austin, Texas, US.

Da Silva, S. P., and Ruwanpura, J. (2006). "Concreting productivity improvements in Alberta on commercial construction projects." *Proc., International Specialty Conference,* CSCE, Calgary, Alberta, Canada, CT-029, 1–8.

Dissanayake, M., Fayek, A. Robinson, Russell, A. D., and Pedrycz, W. (2005). “A hybrid neural network for predicting construction labour productivity.” *Proc., Intern. Computing in Civil Eng.,* ASCE*,* Cancun, Mexico, 1–12.

Dozzi, S.P., and AbouRizk, S.M. (1993). “Productivity in construction.” *Institute for Research in Construction*, National Research Council, Ottawa, Ontario, Canada.

Drewin, F.J. (1982). *Construction productivity: measurement and improvement through work study,*1st Ed., Elsevier Science Publishing Inc.

Durdyev, S., and Mbachu, J. (2011). "On-site labour productivity of New Zealand construction industry: Key constraints and improvement measures." *Aust. J. Constr. Econ. Building*, 11(3), 18–33.

El-Gohary, K.M., and Aziz, R.F. (2014). “Factors influencing construction labor productivity in Egypt.” *J. Manage. Eng.*, 30(1), 1–9.

El-Rayes, K., and Mosehli, O. (2001). “Impact of rainfall on the productivity of highway construction.” *J. Constr. Eng. Manage.,* 127(2), 125–131.

Eslamdoost, E., and Heravi, G. (2013). “Identification and evaluation of effective factors on labor productivity in power plant construction projects.” *Proc. Construction Specialty Conference*, CSCE, Montreal, Quebec, Canada, 10 pp.

Ezeldin, A. S., and Sharara, L. M. (2006). "Neural networks for estimating the productivity of concreting activities." *J. Constr. Eng. Manage*., 132(6), 650–656.

Fayek, A. Robinson. (2011). *Fuzzy hybrid modeling for strategic construction analysis and delivery*. Natural Sciences and Engineering Research Council of Canada Industrial Research Chair proposal, Dept. of Civil and Environmental Engineering, Univ. of Alberta, Edmonton, Alberta, Canada.

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

Fayek, A. Robinson, Shaheen, A., and Oduba, A. (2002). *A pilot study to examine the effective integration of apprentices into the industrial construction sector*, Dept. of Civil and Environmental Engineering, Univ. of Alberta. Presented to the Workforce Development Coordinating Committee, Construction Owners Association of Alberta, Edmonton, Alberta, Canada.

Gerek, İ. H., Erdis, E., Mistikoglu, G., and Usmen, M. (2014). “Modelling masonry crew productivity using two artificial neural network techniques.” *J. Civ. Eng. Manage.*, (ahead-of-print), 8pp.

Gouett, M. C., Haas, C. T., Goodrum, P. M., and Caldas, C. H. (2011). “Activity analysis for direct-work rate improvement in construction.” *J. Constr. Eng. Manage*., *137*(12), 1117–1124.

Grau, D., Caldas, C. H., Haas, C.T., Goodrum,P. M., and Gong, J. (2009). “Assessing the impact of materials tracking technologies on construction craft productivity.” *Autom. Constr.,* 18(7): 903–911.

Gray, A.R., and MacDonell, S.G. (1997). “A comparison of techniques for developing predictive models of software metrics.” *J. Inf. and Software Tec.*, 6(39), 425–437.

Halpin, D. W. (2007). “Fifty years of progress in construction engineering research.” *J. Constr. Eng. Manage.,* 133(9), 635-639.

Handa, V.K., and Abdalla, O. (1989). “Forecasting productivity by work sampling.” *J. Cons. Mgmt. Econ*., 7, 19–28.

Hanna, A. S., (2010). *Construction labour productivity management and methods improvement*. n.p., Madison, WI.

Hanna, A. S., Chang, C.K., Sullivan, K.T., and Lackney, J.A. (2008). "Impact of shift work on labor productivity for labor intensive contractor." *J. Constr. Eng. Manage*., 134(3), 197–204.

Hanna, A. S., Russel, J., Gotzion, T., and Nordheim, E. (1999a). “Impact of change orders on labor efficiency for mechanical construction.” *J.Constr. Eng. Manage.*, 125(3), 176–184.

Hanna, A. S., Russel, J., Nordheim, E., and Bruggink, M. (1999b). “Impact of change orders on labor efficiency for electrical construction.” *J.Constr. Eng. Manage.*, 125(4), 224–232.

Hanna, A. S., Taylor, C. S., and Sullivan, K. T. (2005). "Impact of extended overtime on construction labor productivity." *J. Constr. Eng. Manage*., 131(6), 734–739.

Hendrickson, C., Martinelli, D., and Rehak,D. (1987). "Hierarchical rule-based activity duration estimation," *J. Constr. Eng. Manage*., 113 (2), 288–301.

Heravi, G., and Eslamdoost, E. (2015). "Applying artificial neural networks for measuring and predicting construction-labor productivity." *J. Constr. Eng. Manage*., 04015032.

Herbsman, Z., and Ellis, R. (1990). “Research of factors influencing construction productivity.” *Constr. Manage. Econ*., 8(1), 49–61.

Horman, M. J., and Kenley, R. (2005). "Quantifying levels of wasted time in construction with meta-analysis." *J. Constr. Eng. Manage*.,131 (1), 52–61.

Huh, Y. 2004. “Highway bridge construction production rates for time estimation.” PhD thesis, Univ. of Texas at Austin, Austin, Texas, US.

Ibbs, C. (1997). “Quantitative impacts of project change: Size issues.” *J. Constr. Eng. Manage.*, 123(3), 308–311.

IMF (2015). “World economic outlook.” International Monetary Fund, < http://www.imf.org/external/pubs/ft/weo/2015/01/ > (June 28, 2015).

Jang, H., Kim, K., Kim, J., and Kim, J. (2011). “Labour productivity model for reinforced concrete construction projects.” *J. Constr. Innov.: Info. Proc. Manage.*, 11(1), 92–113.

Jarkas, A., and Bitar, C. (2012). “Factors affecting construction labor productivity in Kuwait.” *J. Constr. Eng. Manage.*, 138(7), 811–820.

Jarkas, A., and Horner, M. (2011). "Revisiting the applicability of learning curve theory to formwork labour productivity." *Constr. Manage. Econ.,* 29(5), 483–493.

Jayawardane, A. K. W., Price, A. D., and Harris, F. C. (1995). “Measurement of brickwork and blockwork productivity: Part A.” *Building Res. Inform.*, 23(2), 81–86.

Jergeas, G. (2009). “Improving construction productivity on Alberta oil and gas capital projects.” Alberta Finance and Enterprise, Alberta, Canada.

Josephson, P., and Björkman, L. (2013). “Why do work sampling studies in construction? The case of plumbing work in Scandinavia.” *Eng. Construct. Architect. Manage.*, 20(6), 589–603.

Koehn, E., and Brown, G. (1985). “Climatic effects on construction.” *J. Constr. Eng. Manage.*, 111(2), 129–137.

Knight, K., and Fayek, A. Robinson (2000). “A preliminary study of the factors affecting the cost escalation of construction projects.” *Can. J. Civ. Eng.*, 27(1), 73–83.

Knowles, R. P. (1997). “Predicting construction labour productivity using neural networks.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Liberda, M., Ruwanpura, J., and Jergeas, G. (2003). "Construction productivity improvement: A study of human, management and external issues." *Proc., Construction Research Congress*, ASCE, Honolulu, Hawaii, 1–8.

Lin, C., and Huang, H., 2010. “Improved baseline productivity analysis technique.” *J. Constr. Eng. Manage*., 136(3), 367–376.

Lim, E. C., and Alum, J. (1995). "Construction productivity: issues encountered by contractors in Singapore." *Inter. J. Proj. Manage*., 13(1), 51–58.

Liou, F., and Borcherding, J. D. (1986). "Work sampling can predict unit rate productivity." *J. Constr. Eng. Manage*., 112(1), 90–103.

Lu, M. (2001). “Productivity studies using advanced ANN models.” PhD thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Lu, M., AbouRizk, S., and Hermann, U. H. (2000). "Estimating labor productivity using probability inference neural network." *J. Comput. Civ. Eng.,* 14(4), 241–248.

Mao, H. (1999). “Estimating labour productivity using fuzzy set theory.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Moselhi, O., Aly, R., and Hassanein, A. (2011). “Labour productivity in building construction: a field study.” *Proc., 3rd International / 9th Construction Specialty Conference*, Ottawa, Canada.

Mosehli, O., Assem, I., and El-Rayes, K. (2005). Change orders impact on labor productivity.” *J. Constr. Eng. Manage.,* 131(3), 354–359.

Moselhi, O., Charles, L., and Fazio, P. (1991). “Impact of change orders on construction productivity.” *Can. J. Civ. Eng.*, 18(4), 484–492.

Moselhi, O., and Khan, Z. (2012). “Significance ranking of parameters impacting construction labor productivity.” *J. Constr. Innov.: Info. Proc. Manage.*, 12(3), 272–296.

Nasirzadeh, F., and Nojedehi, P. (2013). "Dynamic modeling of labor productivity in construction projects." *Int.J.Project Manage.,* 31(6), 903–911.

Oduba, A. O. (2002). “Predicting industrial construction productivity using fuzzy expert systems.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

O’Connor, J. T., and Huh, Y. (2006). “Crew production rates for contract time estimation: Beam erection, deck and rail of highway bridges.” *J. Constr. Eng. Manage*., 132(4), 408–415.

Oglesby, C. H., Parker, H. W., and Gregory, A. H. (1988). *Productivity improvement in construction*. McGraw-Hill, New York, New York, US.

Oral, E. L., and Oral, M. (2010). "Predicting construction crew productivity by using self organizing maps." *Autom. Constr.,* 19(6), 791–797.

Oral, M., Oral, E. L., and Aydın, A. (2012). "Supervised vs. unsupervised learning for construction crew productivity prediction." *Autom. Constr.,* 22, 271–276.

Panas, A., and Pantouvakis, J. P. (2010). “Evaluating research methodology in construction productivity studies.” *Built Hum. Env. Rev.,* 3(1), 63–85.

Park, H.S. (2002). “Development of a construction productivity metrics system (CPMS).” Ph.D. thesis, Univ. of Texas, Austin, Texas, US.

Picard, H.E. (2004). “Driving down construction project labour cost,” *Proceeding, Structural and Construction Engineering*, Rome, Italy.

PMI (Project Management Institute). (2007). *Construction extension to the PMBOK (Project management body of knowledge) guide*. Newtown Square, Pennsylvania, US.

Portas, J., and AbouRizk, S. (1997). "Neural network model for estimating construction productivity." *J. Constr. Eng. Manage.*,123(4), 399–410.

Richard, F., and Liu, A. (2008). *Research methods for construction.* 3rd Ed., Blackwell Publishing Ltd, West Sussex, U.K.

Rivas, R. A., Borcherding, J. D., Gonzalez, V., and Alarcon, L. F. (2011). "Analysis of factors influencing productivity using craftsmen questionnaires: Case study in a Chilean construction company." *J. Constr. Eng. Manage.*,137(4), 312–320.

Sandbhor, S., and Botre, R. (2014). "Applying total interpretive structural modeling to study factors affecting construction labour productivity." *Austr. J. Constr. Econ. Build.,* 14(1), 20–31.

Sanders, S. R., and Thomas, H. R. (1993). "Masonry productivity forecasting model." *J. Constr. Eng. Manage.*, 119(1), 163–179.

Shahtaheri, M., Nasir, H., and Haas, C. T. (2014). "Setting baseline rates for on-site work categories in the construction industry." *J.Constr.Eng.Manage.,*141(5), 04014097.

Sheppard, B., and Lewicki, R. J. (1987). “Toward general principles of managerial fairness.” *Soc. Just. Res.*, 1(2), 161–176.

Silva, S. P., and Ruwanpura, J. Y. (2006). “Concreting productivity improvements in Alberta on commercial construction projects.” *Proc., 1st Int. Constr. Specialty Conf.*, Canadian Society for Civil Engineering, Calgary, Alberta, Canada.

Song, L., and AbouRizk, S. M. (2008). "Measuring and modeling labor productivity using historical data." *J. Constr. Eng. Manage*., 134(10), 786–794.

Sonmez, R., and Rowings, J. E. (1998). "Construction labor productivity modeling with neural networks." *J. Constr. Eng. Manage*., 124(6), 498–504.

Sonmez, R. (1996). “Construction labor productivity modeling with neural networks and regression analysis.” PhD thesis, Iowa State Univ., Ames, Iowa, US.

Srinavin, K., and Mohamed, S. (2003). "Thermal environment and construction workers’ productivity: some evidence from Thailand." *Build. Environ.,* 38(2), 339–345.

Thomas, H. R. (1991). "Labor productivity and work sampling: the bottom line." *J. Constr. Eng. Manage*., 117(3), 423–444.

Thomas, H. R., Guevara, J. M., and Gustenhoven, C. T. (1984). “Improving productivity estimates by work sampling.” *J. Constr. Eng. Manage*., 110 (2), 178–188.

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). "Modeling construction labor productivity." *J. Constr. Eng. Manage*., 116(4), 705–726.

Thomas, H. R., Mathews, C. T. (1986). “Learning curve models of construction productivity” *J. Constr. Eng. Manage*., 112 (2), 245–258.

Thomas, H. R., and Raynar, K. A. (1997). "Scheduled overtime and labor productivity: quantitative analysis." *J. Constr. Eng. Manage*., 123(2), 181–188.

Thomas, H. R., and Sakarcan, A. S. (1994). "Forecasting labor productivity using factor model." *J. Constr. Eng. Manage*., 120(1), 228–239.

Thomas, H., and Yiakoumis, I. (1987). “Factor model of construction productivity.” *J. Constr. Eng. Manage.,* 113(4), 623–639.

Tsehayae, A. A., and Fayek, A. Robinson. (2012). “A research framework for work sampling and its application in developing comparative direct and support activity proportions for different trades,” *Proc., Construction Research Congress*, ASCE, West Lafayette, Indiana, US, 1420–1429.

Tsehayae, A. A., and Fayek, A. Robinson. (2014a). "Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects." *Can. J. Civ. Eng.,* 41(10), 878–891.

Tsehayae, A.A., and Fayek, A. R. (2014b). *Data collection protocol: Labour productivity analysis and modeling*. Technical report submitted to Suncor Energy, NSERC IRC SCMD: Productivity/Suncor/ Data Collection Protocol Report/2014–TD02, Edmonton, Alberta, Canada, October 14, 2014, 99 pp.

Watkins, M., Mukherjee, A., Onder, N., and Mattila, K. (2009). "Using agent-based modeling to study construction labor productivity as an emergent property of individual and crew interactions." *J. Constr. Eng. Manage*., 135(7), 657–667.

Xue, X., Shen, Q., Wang, Y., and Lu, J. (2008). “Measuring the productivity of the construction industry in China by using DEA-based Malmquist productivity indices.” *J. Constr. Eng. Manage*. 134(1), 64–71.

Yi, W., and Chan, A. (2014). "Critical Review of Labor Productivity Research in Construction Journals." *J. Manage. Eng.,* 30(2), 214–225.

Chapter 3: Identification of Key Input Parameters Influencing CLP[3]

3.1: INTRODUCTION

The input parameters, which are made up of various objective and subjective factors and practices (e.g., crew size, foreman skill in planning, complexity of task, quality of drawings, weather condition etc.), define the internal and external environments that in turn, whether directly or indirectly, influence the efficiency of the conversion process. The parameters are also used to understand the cause and effect relation of the environment to the efficiency of the conversion process, represented in terms of efficiency measures like CLP. However, CLP is situated in an environment that is more complex and unpredictable than the conversion process itself. Accordingly, several CLP studies have identified numerous input parameters defining the internal and external environments of CLP (Oduba 2002; Liberda et al. 2003; CII 2006; Song and AbouRizk 2008; Tsehayae and Fayek 2014b; Gerek et al. 2014). However, the identified influencing parameters and the associated CLP models were context-dependent, as the identified parameters and their degree of impact on CLP varied from project to project (Gerek et al. 2014). Context plays an active part in CLP research analysis as it is invariably dynamic and imperative for the development of meaningful findings (Engwall 2003). However, only a few CLP studies explicitly defined the context of the CLP study.

Given the often large project development expenditures made by construction owner and execution organizations, the project team's ability to understand the project context and accurately predict CLP is a key element in the analysis and control of project costs. As skilled construction labour is a scarce resource (Dai and Goodrum 2012), its effective use remains a priority; optimizing CLP through appropriate analysis and modeling is therefore critical. Such analysis and modeling requires identification of the multilevel, complex, and context-dependent key parameters (factors and practices) influencing CLP in different project contexts. However, despite the extensive research in the area, consensus on the identification and generalization of key influencing parameters is yet to be achieved (Panas and

---

[3] Parts of this chapter have been published in Canadian Journal of Civil Engineering, Volume 41, Issue 10, pp. 878-891, and the Proceedings, ASCE Construction Research Congress 2014, Atlanta, Georgia, US, May 19-21, pp. 837-846.

Pantouvakis 2010). The identification of the key influencing parameters also reduces the large feature space, resulting from the numerous input parameters, thereby improving the interpretability and accuracy of CLP models. In a study of concrete pouring, formwork, and concrete finishing activities, CLP models having fewer significant parameters showed better prediction than models based on many parameters which did not consider their significance (Sonmez and Rowings 1998). Gerek et al. (2014) also showed that removing some parameters improved the accuracy of masonry activity CLP model.

The reduction of the input feature space through identification of the key influencing input parameters can be carried out using either an expert or data-driven approaches. Expert-driven approaches focus on factor surveys for collecting expert opinions, and identify the key parameters based on perceived influence on CLP. On the other hand, data-driven techniques utilize actual data of input parameters and CLP, and identify the key parameters based on their relation with CLP. In this research, both approaches are employed. First, the expert-driven approach is used to: (1) verify the hierarchal list of parameters influencing CLP, (2) establish properties of the parameters by identifying their level of existence together with the most appropriate project member to be targeted, so as to gather accurate parameter data values, and (3) establish whether key parameters influencing CLP are context-specific. Then, based on the findings of the expert-driven approach, data was collected on the parameters and key parameters were identified using the data-driven approach.

In this chapter, the methodology for identifying key input parameters (factors and practices) influencing labour productivity using expert and data-driven approaches is presented and discussed. To begin with, a literature review of studies that established key parameters for the analysis and modeling of CLP is presented. Then, the details of the expert-driven approach methodology, data collection, and analysis is discussed. Next, using the data collected from 141 surveys comprehensively addressing project management (PM) and trade respondents from six Canadian projects, the expert-driven methodology and findings on properties of the parameters and identified key parameters are presented and discussed. The chapter also explores whether a gap exists in experts' (project management and trade groups') knowledge in identifying the critical parameters. Finally, the data-driven approach for identifying the key parameters using feature selection algorithm and the collected field data is discussed.

3.2: BACKGROUND AND LITERATURE REVIEW

Because of its significance to project performance, CLP has been well studied. As discussed in Section 2.2.1 and summarized in Appendix A.1, numerous input parameters that influence CLP have been identified. Most of the past studies identifying key input parameters used an expert-driven approach. Lim and Alum (1995) identified 17 influencing factors grouped under three categories and carried out surveys with 67 respondents from civil engineering and building contractors in Singapore to identify the relative importance of the identified 17 factors and establish the top six critical factors. Liberda et al. (2003) identified 51 factors grouped under three categories—human, external, and management—and carried out an interview survey with 20 project management experts from Alberta, Canada to identify the relative importance of the identified 51 factors and establish the top 15 critical factors. CII (2006) identified 83 influencing factors grouped under 11 categories: supervisor direction, communication, safety, tools and consumables, materials, engineering drawing management, labour, foreman, superintendent, project management, and construction equipment. In the most extensive and detailed existing study of factors influencing CLP, Dai et al. (2009) carried out a factor survey based on an identified 83 factors with 1,996 craftspeople on 28 U.S. industrial construction projects to identify the top 10 critical factors and the relationship between these factors. Detailed investigation of differences in perception between the different generations of craft workers (Dai and Goodrum 2012), foremen and craft workers (Dai et al. 2007), union and non-union craft workers, and trades (Dai et al. 2009) was carried out using the same survey data; significant differences were observed between the compared groups.

Durdyev and Mbachu (2011) identified 56 influencing factors grouped under two main categories and carried out surveys with 37 respondents from project management consultants, contractors, and subcontractor in New Zealand to identify the relative importance of the identified 56 factors and establish the top eight critical factors. Jarkas and Bitar (2012) identified 45 influencing factors grouped under four categories and conducted a questionnaire survey with 157 respondents from civil engineering and buildings construction firms in Kuwait to identify the relative importance of the identified 45 factors and establish the top five critical factors. Eslamdoost and Heravi (2013) identified 15 factors grouped under five categories and carried out a survey with 106 site and office staff in thermal power plant construction projects in Iran; they were able to establish the relative importance of the identified 15 factors and five factor categories. Table 3.1 shows a summary of previous studies that identified key input parameters. The list of the identified key input parameters is ranked in descending order of influence on CLP.

Table 3.1: Literature Summary of Key Input Parameters Influencing CLP

Study details	**Key parameters ranked in descending order of influence**
Lim and Alum (1995) 6 Key Factors	Difficulty in recruitment supervisors; Difficulty in recruiting workers; High rate of labour turnover; Absenteeism at work site; Communication problems with foreign workers; Inclement weather that requires work stoppage for one day or more
Liberda et al. (2003) 15 Key Factors	Lack of detailed planning; Worker experience and skills; Inadequate supervision; Worker motivation; Non availability of materials; Worker attitude and morale; Team-spirit of the crew; Non availability of information; Changes in drawings and specifications; Non availability of tools; Non availability of equipment; Nature of project (size and complexity); Lack of procedures for construction methods; Changes in contract; Congested work area
Dai et al. (2009) 10 Key Factors	I have to wait for people and/or equipment to mode the material I need; There are errors in the drawings that I use; When there is a question or problem with a drawing, the engineers are slow to address the issue; If I need a manlift to do my job, there aren't any available; When I need a crane or forklift to help me, there aren't any available; I can't get the consumables I need to do my job; I have to search in a lot of places to find the tools I need to do my job; When I go to install prefabricated items, work has to be done on them to fix quality problems; I can't get the power tools from the contractor that I need to do my job; My supervisor does not provide me with enough information to do my job
Durdyev and Mbachu (2011) 8 Key Factors	Reworks; Level of skill and experience of the workforce; Adequacy of method of construction; Buildability issues; Inadequate supervision and coordination; Statutory compliance; Unforeseen events; Wider external dynamics
Moselhi and Khan (2012) 9 Factors	Temperature; Work type; Floor level; Wind speed; Labour percent; Precipitation; Gang size; Humidity; Work method
Jarkas and Bitar (2012) 5 Key Factors	Clarity of technical specifications; extent of variation/change orders during execution; coordination level among various design disciplines; lack of labour supervision; proportion of work subcontracted

However, the use of data-driven approach for identifying key input parameters has been limited due to the lack of field data on input parameters and CLP. As discussed in Section 2.3, the parameters affecting labour productivity are numerous, complex, interlinked, dynamic, and involve a mix of subjective and objective concepts, thus making field data collection a challenging task. In rare cases where detailed parameter documentation was carried out together with documentation of the output parameter (achieved labour productivity), data-driven methods have been employed to identify critical parameters. In a recent publication, Moselhi and Khan (2012) compared three data-driven CLP parameter ranking approaches, namely, fuzzy subtractive clustering, neural networks, and stepwise variable selection procedure, for evaluating the influence of nine parameters on labour productivity of formwork installation activity. Using a ten month field data from two building construction projects, located in Montreal, Canada; the authors showed that the three most important parameters were identified in the same order by the artificial intelligence based methods (fuzzy subtractive clustering and neural networks); however the step wise regression analysis provided somewhat different results.

Despite extensive research in the area, consensus on the development of context-specific key parameters and generalization of common parameters is yet to be achieved (Tsehayae and Fayek 2014a; Tsehayae and Fayek 2014b). Past studies ranked factors irrespective of positive or negative influence and mainly focused on factors negatively influencing CLP, while failing to comprehensively address the factors positively influencing CLP. They considered the perspective of either PM or trade level respondents without addressing differences in perspective between the two groups, thereby misrepresenting the multilevel and context-specific nature of productivity factors. Additionally, researches mainly identified critical parameters using expert-driven approach, which were based on a limited number of parameters and expert opinions; this approach has not provided a better understanding of the parameters and their impact on labour productivity (Panas and Pantouvakis 2010). Furthermore, past studies did not investigate practices as key factors influencing CLP. Therefore, in this chapter, input parameters made up of factors and practices are investigated using both expert and data-driven approaches; and the study methodology and findings presented in this chapter attempt to address the aforementioned gaps in previous research.

3.3: EXPERT-DRIVEN APPROACH FOR IDENTIFICATION OF KEY INPUT PARAMETERS OF CLP

3.3.1: Research Methodology of Expert-driven Approach

The expert-driven approach uses surveys to identify the key parameters (factors and practices) positively and negatively influencing CLP according to project management (PM) and trade level respondents from a number of ongoing construction projects. The list of 169 parameters influencing CLP categorized into a hierarchal structure as shown in Table 2.1 was used to develop the surveys for collecting expert opinions. While CLP tends to be considered a micro level subject wherein a group of organized workers are required to transform a set of inputs into project outputs, parameters influencing CLP are multilevel, ranging from macro, to meso, and to micro levels. Accordingly, for this study, macro parameters are defined to include organizational, provincial, national, and global level parameters; meso parameters are defined to include project level parameters; and micro parameters are defined to include task-at-hand level parameters directly related to the on-site workforce.

The survey, administered to personnel from various organizational levels of participating construction companies executing the projects under study, required a systematic approach to address the key factors and practices influencing CLP. The survey firstly addressed the various levels of factors and practices—from micro and some meso parameters at the trade level (craftspeople and foremen) to meso and macro parameters and practices at the PM level (Dai et al. 2009)—and their effects on CLP. Two surveys were developed to address parameters relevant to the trade (craftspeople and foremen) and PM (project managers, supervisors, and superintendents) levels. Secondly, the survey also addressed the differences in worker perspectives between project levels by collecting these perspectives in terms of agreement on the rankings of the identical micro and meso parameters included in both surveys. The perspective aspect provides a better understanding of the individual parameters and their relevance for further study, and also help in identifying the most appropriate project member for gathering parameter data. The two survey forms, namely the project management survey (PM survey) and the trade survey, were thus developed to meet the two design objectives. The two surveys together with respondent information letters and consent form were developed for data collection (Tsehayae and Fayek 2014c).

The PM survey addresses some micro (activity), all meso (project), and all macro (organizational, provincial, national, and global) level parameters—a total of 141 parameters in 16 categories, as shown in Table 3.2. The trade survey addresses all micro (activity) and some meso (project) level parameters—a total of 89 parameters in 9 categories, as shown in Table 3.2. All the identified practices (PM practices and project best practices) were included in the PM survey.

Table 3.2: Factor Surveys: Categories of Parameters

Category label	**Parameter category**	**Number of parameters**		
		PM survey	**Trade survey**	**Common**
A	Labour and crew	18	27	10
B	Material and consumables	6	9	5
C	Equipment and tools	4	8	3
D	Foreman	6	9	3
E	Task property	3	7	3
F	Location property	5	10	1
G	Project delivery and contract	2	*	*
H	Engineering and instructions	2	4	2
I	Project complexity	2	*	*
J	Health, safety, and environment	11	4	2
K	Project management practices	19	*	*
L	Project best practices	14	*	*
M	Project owner nature	9	*	*
N	Management of project	*	11	*
O	Organizational	9	*	*
P	Provincial	13	*	*
Q	National	11	*	*
R	Global	7	*	*
	Total	**141**	**89**	**29**

Note: *Denotes that the parameters category is not included in the survey

Both surveys collected respondents' opinions on parameters influencing labour productivity in the given context under study—unlike past surveys, which focused on respondents' general or context-free opinions on parameters influencing CLP. As the parameters influencing CLP are multilevel, it is important to determine instances where PM and trade personnel disagree. While a higher level of agreement on parameters between the two groups will help in implementing improvement strategies, a lack of agreement will demand further investigation into the sources of these differences of perspective. To investigate these differences between respondent groups, a total of 37 parameters in 9 categories, illustrated in Table 3.2, are common to the PM and trade surveys.

The surveys were first distributed to a focus group representing different construction companies, and were then improved and revised according to the feedback received. The surveys also allowed for focus group respondents to add parameters they considered important, and necessary revisions were made, and the final 169 parameter were verified. The surveys have a background section to collect the general attributes of the respondents in terms of demographic information, union status, trade, and position. Survey questions are divided into two sections: (1) agreement or frequency, and (2) impact. The agreement or frequency section evaluates the extent to which each parameter exists in the given project setting. Agreement questions, as shown in Table 3.3, are for parameters that will become an issue if they occur on a continual basis (e.g., crew size is not adequate for the task at hand). Frequency questions focus on parameters that occur with varying frequencies (e.g., power equipment breakdown). Frequency questions, as shown in Table 3.3, are only used for micro level parameters (activity factors related to material and equipment).

Table 3.3: Sample Agreement/Frequency and Impact Survey Statements

Parameter	Agreement							Impact						
	Strongly Disagree	Disagree	Slightly Disagree	No Opinion	Slightly Agree	Agree	Strongly Agree	Strongly Negative	Negative	Slightly Negative	No Impact	Slightly Positive	Positive	Strongly Positive
The work area is clean	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Parameter	Frequency							Impact						
	Never	Very Rarely	Rarely	Sometimes	Often	Very Often	Constantly	Strongly Negative	Negative	Slightly Negative	No Impact	Slightly Positive	Positive	Strongly Positive
I wait in a line for manlifts	1	2	3	4	5	6	7	1	2	3	4	5	6	7

All the parameters common to the PM and trade surveys were designed as agreement-type survey statements. All questions also assess parameter influence (positive, negative, or none at all) on CLP for the project under study. Past studies have mainly relied on the possible negative influencers of CLP (Dai et al. 2009; Jarkas and Bitar 2012) and have not properly addressed the positive influencers. This study aims to identify both kinds of influencers, so as to use the positive influencing parameters to further improve work conditions while addressing the undesirable effects of the negatively influencing parameters.

The survey is similar in structure to the "Voice of the Worker" survey (CII 2006; Dai et al. 2009): bipolar, seven-point Likert scales structured into positively and negatively worded statements collect ratings on agreement with/frequency of and impact of parameters, and the two scales then enable the analysis and ranking. In the administered survey, the parameters are presented in both positively and negatively worded statements, as shown in Table 3.3, in order to improve the accuracy of responses by ensuring respondents pay attention to each parameter statement (Stewart and Frye 2004). Once the required value for each statement has been determined by individual respondents, calculations for positive and negative effect scores, as shown in the next section, are performed to determine the rankings for the positive and negative effects of the various parameters.

3.3.1.1: Parameter Survey Collection

As discussed in Section 2.5, in order to identify the participating construction companies, an invitation was first sent to a number of construction companies in the province of Alberta, Canada and a workshop was held to introduce the research and its data collection protocol. A total of seven companies involved in commercial (three), residential (two), and industrial (two) projects agreed to participate in the study and provided a total of six projects, as shown in Table 3.4, from which data used to identify the key parameters of the given contexts were collected.

Table 3.4: Profiles of Companies and Projects

Characteristics	Context	Type	Number of projects	Notes
Companies (7)	Building (5)	Commercial	3	1 Main contractor and 2 subcontractors
		Residential	2	2 Main contractors
	Industrial (2)	Turnaround	2	2 Main contractors
Projects (6)	Building (3)	Commercial	1	Commercial mixed-use office-staff facility building
		Residential	2	Seniors residence; Mixed apartment-community center building
	Industrial (3)	Turnaround	3	Coal power plant boiler shutdowns

Note: Values in bracket represent the total values in the group.

The projects under study have been divided into two main context categories based on industry type. The first deals with the building construction context and involves commercial and residential projects; the second deals with the industrial construction context and involves industrial plant shutdown projects. Accordingly, since there was only one project in the commercial category, the survey data from that project were merged with that of the residential projects and the context was classified as the building project context.

The data collection effort produced a total of 141 surveys: 78 for the building context and 63 for the industrial context. Out of the 141 surveys, 42 were from PM staff (project manager, superintendent, estimator, coordinators, etc.) and 99 were from trade staff (foremen, apprentices, helpers, and others). Details of the participating companies, characteristics of the projects, and respondents are shown in Table 3.4. Surveys were administered to project personnel from the participating construction companies. Determining sample size—the number of respondents to be surveyed from the population of workers—was essential to ensure the reliability and accuracy of results. As the survey addresses parameters from macro to micro levels, respondents from different levels of the project were sought, and the population (number of workers in a given project) for the survey was assumed to be comprised of all construction project-related personnel for each of the projects under study. This study population ensures that the critical parameters identified through the survey will be applicable to each company's context and its project work force. The survey population in terms of the total workers was stratified into PM and trade levels. Once the population for each stratum was established, random sampling was taken. The stratified random sampling technique was considered an appropriate method in this situation, as the structure within the population of each stratum is assumed to be similar in terms of role and function, and an adequate sample size is used to ensure proper representation of the population as a whole. The sampling aim in this study was to achieve a 10% margin of error and 95% confidence interval based on the established populations as shown in Table 3.5. Using the Cochran formula Eq.(3.1) for determining survey sample size and for a 10% margin of error and 95% confidence interval, the sample size for the infinite population (SS_{ip}) is 96 (Jarkas and Bitar 2012). Thus, the target sample is 32 for the PM survey and 80 for the trade survey, based on the study population of 46 for the PM survey and 459 for the trade survey as shown in Table 3.5. Table 3.5 shows that of the collected surveys, 42 responses were received for the PM survey and 99 for the trade survey; these numbers fulfill the target size requirements and the sample is therefore deemed adequate for analysis.

$$Survey\ Sample\ Size_{per\ strata} = \frac{SS_{ip}}{1 + \frac{SS_{ip} - 1}{Population\ of\ Strata}} \tag{3.1}$$

Table 3.5: Survey Data Collection and Respondents' Profile per Context

Context	Category	Population	Sampled	Average years of experience	Respondents' profile
Building	PM survey	24	20	11.0	Construction manager (15%), project manager (35%), superintendent (10%), safety supervisor (10%), project coordinator (10%), other (20%)
	Trade survey	59	58	7.9	Carpenters (34%), scaffolders (3%), labourers (5%), crane operator (3%), electricians (40%), mechanical (9%), other (5%)
Industrial	PM survey	22	22	7.4	Construction manager (9%), project manager (5%), project control (27%), superintendent (27%), project coordinator (9%), planner (9%), other (5%)
	Trade survey	400	41	15.5	Boilermakers (27%), sheet metal workers (2%), labourers (2%), welders (15%), millwright (34%), electricians (7%), other (12%)
	Total	**505**	**141**		

Of PM survey respondents in the building context, whose average years of experience was 11 years, 95.0% rated the management of the current project as “Good” or better; of PM survey respondents in the industrial context, whose average years of experience was 7.4 years, 68.2% rated the management of their current project as “Good” or better. The trade survey respondents’ positions were journeyman (51.5%), apprentice (29.3%), foreman (14.1%), helper (4.0%), and other (1.0%). Of the trade respondents in the building context, whose average years of experience was 7.9 years, 87.9% rated the management of the current project as “Good” or better, and of trade respondents in the industrial context, whose average years of experience was 15.5 years, 80.5% rated the management of their current project as “Good” or better. Overall, the respondents represented a broad range of positions and trades within each project, reported adequate industry experience, and rated the management of their projects as good or better.

3.3.1.2: Survey Data Analysis to Determine Parameter Evaluation Scores

The analysis in this study expands upon similar work by the CII (2006) and Dai et al. (2009) by not only addressing the ranking of parameters, but also by exploring both their positive and negative effects on productivity. Furthermore, the current survey design enables comparison of the PM and trade perspectives, furthering understanding of the parameters and their possible use in CLP improvement strategies.

For agreement-type parameters, which are presented in either positive or negative parameter statements, first, the level of agreement (R_A) or disagreement (R_D) with a given parameter statement by a number of respondents was computed using equations (3.2) and (3.3), where the maximum possible weighted percentage of agreement or disagreement is equal to 50:

$$R_A = \frac{(A\ x\ 1 + B\ x\ 2 + C\ x\ 3)}{6}\ x\ 100 \tag{3.2}$$

$$R_D = \frac{(D\ x\ 1 + E\ x\ 2 + F\ x\ 3)}{6}\ x\ 100 \tag{3.3}$$

where A = percentage of respondents rating the positively worded parameter as 5 (slightly agree); B = percentage of respondents rating the positively worded parameter as 6 (agree); C =

percentage of respondents rating the positively worded parameter as 7 (strongly agree); D = percentage of respondents rating the negatively worded parameter as 3 (slightly disagree; E = percentage of respondents rating the negatively worded parameter as 2 (disagree); and F = percentage of respondents rating the negatively worded parameter as 1 (strongly disagree).

The impact in terms of positive (I_P) or negative (I_N) impact of a given agreement-type parameter statement by a number of respondents was computed using equations (3.4) and (3.5), where the maximum possible weighted percentage of positive or negative impact is equal to 50:

$$I_P = \frac{(X \ x \ 1 + Y \ x \ 2 + Z \ x \ 3)}{6} \ x \ 100 \tag{3.4}$$

$$I_N = \frac{(U \ x \ 1 + V \ x \ 2 + W \ x \ 3)}{6} \ x \ 100 \tag{3.5}$$

where X = percentage of respondents rating the impact of the parameter as 5 (slightly positive); Y = percentage of respondents rating the impact of the parameter as 6 (positive); Z = percentage of respondents rating the impact of the parameter as 7 (strongly positive); U = percentage of respondents rating the impact of the parameter as 3 (slightly negative); V = percentage of respondents rating the impact of the parameter as 2 (negative); and W = percentage of respondents rating the impact of the parameter as 1 (strongly negative). Next, the positive and negative effects of each parameter were evaluated separately. For the positive effect of a positively worded parameter the evaluation index and evaluation score was computed using equations (3.6) and (3.7); similarly, equations (3.8) and (3.9) were used to calculate the positive effect of a negatively worded parameter. First, the evaluation index based on the product of the agreement/disagreement and impact scores was computed. Then, the evaluation score was computed by dividing the evaluation index of a given parameter by the maximum possible evaluation score. The maximum possible evaluation score is equal to 2500—the product of the maximum values of agreement/disagreement (50) and impact (50).

Positively worded parameters and positive effect:

$$Evaluation \ Index_{AP(+ve)} = R_A \ x \ I_P \tag{3.6}$$

$$Evaluation\ Score_{AP(+ve)} = \frac{Evaluation\ Index_{AP(+ve)}}{2500}\ x\ 100 \quad (3.7)$$

Negatively worded parameters and positive effect:

$$Evaluation\ Index_{AN(+ve)} = R_D\ x\ I_P \quad (3.8)$$

$$Evaluation\ Score_{AN(+ve)} = \frac{Evaluation\ Index_{AN(+ve)}}{2500}\ x\ 100 \quad (3.9)$$

For the negative effect of a positively worded parameter, the evaluation index and evaluation score were computed using equations (3.10) and (3.11); similarly, equations (3.12) and (3.13) were used to calculate the negative effect of a negatively worded parameter.

Positively worded parameters and negative effect:

$$Evaluation\ Index_{AP(-ve)} = R_D\ x\ I_N \quad (3.10)$$

$$Evaluation\ Score_{AP(-ve)} = \frac{Evaluation\ Index_{AP(-ve)}}{2500}\ x\ 100 \quad (3.11)$$

Negatively worded parameters and negative effect:

$$Evaluation\ Index_{AN(-ve)} = R_A\ x\ I_N \quad (3.12)$$

$$Evaluation\ Score_{AN(-ve)} = \frac{Evaluation\ Index_{AN(-ve)}}{2500}\ x\ 100 \quad (3.13)$$

The frequency-type parameters were only used for micro parameters under the “material and consumables” and “equipment and tools” categories for the trade survey (see Table 3.3). The parameters were all negatively worded and the evaluation index and scores were calculated as follows. First, the frequency rating of a parameter by a number of respondents (F_R) was computed using the weighted percentage of respondents rating the parameter’s frequency based on the seven-point Likert scale, as shown in equation (3.14), where the maximum possible weighted frequency rating is equal to 100:

$$F_R = \frac{(H\ x\ 1 + I\ x\ 2 + J\ x\ 3 + K\ x\ 4 + L\ x\ 5 + M\ x\ 6 + N\ x\ 7)}{7}\ x\ 100 \quad (3.14)$$

where H = percentage of respondents rating the parameter frequency as 1 (never); I = percentage of respondents rating the parameter frequency as 2 (very rarely); J = percentage of respondents rating the parameter frequency as 3 (rarely); K = percentage of respondents rating the parameter frequency as 4 (sometimes); L = percentage of respondents rating the parameter frequency as 5 (often); M = percentage of respondents rating the parameter frequency as 6 (very often); N = percentage of respondents rating the parameter frequency as 7 (constantly).

Then, the effect in terms of positive or negative impact of a given frequency parameter-type statement by a number of respondents was computed using equations (3.4) and (3.5). Finally, the evaluation index and score was computed for positive effect of the parameter using equations (3.15) and (3.16), and for negative effect using equation (3.17) and (3.18). The evaluation index was based on the product of frequency rating and impact score, and the evaluation score was computed by dividing the evaluation index of a given parameter by the maximum possible evaluation score. The maximum possible evaluation score is equal to 5000—the product of the maximum values of frequency rating (100) and impact (50).

Positive effect:

$$Evaluation\ Index_{F(+ve)} = F_R\ x\ I_P \tag{3.15}$$

$$Evaluation\ Score_{F(+ve)} = \frac{Evaluation\ Index_{F(+ve)}}{5000}\ x\ 100 \tag{3.16}$$

Negative effect:

$$Evaluation\ Index_{F(-ve)} = F_R\ x\ I_N \tag{3.17}$$

$$Evaluation\ Score_{F(-ve)} = \frac{Evaluation\ Index_{F(+ve)}}{5000}\ x\ 100 \tag{3.18}$$

The ranking considered the evaluation scores of the individual parameters in the PM survey and the trade survey. The evaluation scores of each parameter were normalized on a scale of 0 to 100, where an increased evaluation score indicates greater effect, either positive or negative, on CLP. For ranking of the parameter categories shown in Table 3.2, group ranking scores were determined by taking the average evaluation score of the individual parameters under each category. The group ranking scores were then normalized on a scale of 0 to 100. For each of the two surveys (PM survey and trade survey), a list of parameters ranked according to their positive effects on CLP was produced; similarly, a list of parameters ranked according to their negative effects on CLP was also produced for each of the two surveys. For the purpose of comparing all of the various parameters ranked according to their positive effects in discussion of study results, the original survey parameter statements have been modified so that they all read as positive statements which imply favourable conditions for better CLP. Similarly, for the purpose of comparing the various parameters ranked according to their negative effects, the original survey parameter statements have been modified so that they all read as negative statements which imply unfavourable conditions for better CLP. Additionally, using the normalized category evaluation score, computed by taking the average positive and negative evaluation scores of the parameters under each parameter category, the rankings of the parameter categories have also been determined and reported. However, the category evaluation scores and the ranking will also depend on the number of parameters within the category. As shown in Table 3.2, the number of parameters per category varies, and therefore the category evaluation scores will be skewed towards categories with fewer parameters where any of the parameters have a high evaluation score.

3.3.2: Findings and Discussion on Expert-driven Key Input Parameters Influencing CLP

First, the internal consistency or reliability of the PM and trade surveys was examined using Cronbach's alpha method (Stewart and Frye 2004). Since the evaluation scores shown in equations (3.11), (3.13), (3.16), and (3.17) are based on weighted percentages of all responses, it is not possible to use the evaluation scores to measure the different parameters survey statistical values, since the survey statistical values require individual response values; therefore, the use of the impact rating for each parameter is appropriate for such survey designs (CII 2006; Dai et al. 2009). The impact rating responses of the 141 parameters included in the PM survey and 89 parameters included in the trade survey were extracted from the surveys collected in building and industrial contexts. Then, using IBM SPSS 22® statistical package, the Cronbach's alpha statistical values for the PM and trade surveys were determined. The Cronbach's alpha values for the building context PM and trade surveys and for the industrial context PM and trade surveys were found to be 0.938, 0.956, 0.961, and 0.950, respectively.

The overall Cronbach's alpha value for all surveys was actually higher than the minimum value of 0.70 (Eslamdoost and Heravi 2013), which indicated a strong internal consistency or reliability of the PM and trade surveys. Next, the survey results in terms of overall parameter group rankings, top 10 key parameters for each context, and top overall parameters are presented. Furthermore, an investigation into the differences in perspective on parameters (both factors and practices), based on positive and negative effects on CLP, is presented for each of the two contexts: building and industrial.

3.3.2.1: Parameter Category Ranking

The rankings of the parameter categories, with normalized category evaluation scores based on the results of the PM and trade surveys from both the building and industrial contexts, are shown in Appendix B.1. Based on the PM survey respondents from the building and industrial contexts "equipment and tools" category (e.g., "adequate and quality work tools") was identified as the top ranked category by respondents from both contexts, indicating its commonly perceived positive effect on CLP.

A similar ranking of parameter categories by Dai et al. (2009) also confirmed "tools and consumables" as first and construction "equipment" as fourth among 11 categories. The PM respondents from the building context identified global (e.g., "global economy's uncertainty in facing another slow down") as the top category having negative effects on CLP, while industrial context PM respondents identified "engineering and instruction" (e.g., "drawings and specifications unavailability well ahead of implementation") category. Knight and Fayek (2000) also ranked "insufficient/incomplete drawings" as the top ranked factor causing cost escalation of construction projects in Alberta. The top three categories negatively influencing CLP were found to be similar between the two contexts.

The trade survey respondents from the building context identified "foreman" as the top category having positive effects on CLP, while industrial context trade respondents identified "labour and crew" category. The trade respondents from the building context identified "material and consumables" as the top category having negative effects on CLP, which was ranked second among 11 categories by Dai et al. (2009), while industrial context trade respondents identified "equipment and tools" category. Overall, the top three categories influencing CLP positively and the top two categories influencing CLP negatively were ranked similarly by respondents in the two contexts. The negative effect categories identified by respondents from the industrial context conform to the CII study which also identified “equipment and tools”, “materials”, and “engineering drawings” as the top three ranked parameter categories for severity, or negative influence, on CLP based on trade level surveys (CII 2006; Dai et al. 2009). However, it should be noted that the category rankings are dependent not only on the evaluation scores of the individual parameters but also on the total number of parameters within a category.

3.3.2.2: Key Parameters Influencing CLP in Building and Industrial Contexts

Building Context

The rankings of the top 10 parameters, with evaluation scores, identified by PM and trade survey respondents and based on positive and negative effects on CLP for building projects are shown in Table 3.6. According to the PM respondents, of the 141 parameters included in the PM survey, “adequate and quality work tools” was identified as the top parameter having a positive effect. However, Jarkas and Bitar (2012) found “unavailability of suitable tools” ranked only 34th among 45 factors affecting CLP in Kuwait. The “aging of Canada’s population” was identified as the top parameter having a negative effect, which is in line with the expected retirement of many construction workers and the expected shortage of labour supply in the Canadian construction industry (CSC 2013).

Of the 89 parameters included in the trade survey, respondents identified “job site orientation program for new craftsmen” as the top parameter having a positive effect on CLP and “lack of protection from weather effect” as the top parameter having a negative effect. “Weather conditions” and the “need for protection” are in agreement with the findings reported by Knight and Fayek (2000). The rankings of the top 10 key parameters shown in Table 3.6 provide further insight into parameters influencing building projects.

Table 3.6: Building Context: Top 10 Parameters

Rank	PM survey Parameters	ES	Trade survey Parameters	ES
	Positive influence:		**Positive influence:**	
1	There are adequate and quality work tools	100	For new craftsmen, the job site orientation program is carried out	100
2	The organization has many successful years in industry	87.2	Frequency of accidents and personal injury is low	88.8
3	Daily job hazard assessment system is in place	82.6	Foreman has the required experience	85.9
4	Efforts are taken to reduce turnover of foremen	72.6	There is really good cooperation between craftsmen in a crew	78.1
5	Integration management practices: The process required to ensure that the various elements of the project are properly coordinated is properly implemented	69.5	Crew is given adequate training before commencement	74.5
6	Zero accident techniques are effectively applied	64.0	Craftsmen trust in the skills and judgment of their supervisors	74.1
7	Leadership training is provided to foremen	64.0	Crew is experienced and has the necessary competence	70.1
8	Cost management practices: A reporting system at company level is in place for the identification of cost overruns	63.6	Foreman's management style is participative and motivating	67.5
9	Delivered materials are of high quality	62.2	Work is fairly assigned to the different crews	67.3
10	There are adequate material transportation equipment (cranes, forklifts)	60.5	Temporary electrical service is always provided	66.1
	Negative influence:		**Negative influence:**	
1	Canada's population is aging	100.0	The work area is not protected from weather effect	100.0
2	Global economy still faces uncertainty of facing another slow down	56.1	Materials are not delivered on time to task location	88.6
3	Drawings and specifications are often not complete and require updates	54.4	The materials delivered have quality problems	78.8
4	Natural gas prices (dollar/GJ) are currently low	47.8	Stringent safety rules are negatively affecting productivity	74.9
5	The available supervisors for construction projects in Alberta is not adequate	46.9	The work area is congested	65.0
6	Oil prices (dollar/barrel) are highly volatile	45.0	Correction work due to quality problems of prefabricated products is necessary	64.1
7	There are many competing projects within the province	43.9	Work conditions are compromised by excessive noise, dust and fumes	63.9
8	Crew experience and competence is not meeting expectations	42.0	Electrical power gets disconnected during operation	53.1
9	Craftsmen are not flexible in accommodating task changes	38.0	There is a shortage of good transportation equipment (cranes, forklifts)	49.5
10	Prices for outputs (project completion costs) are substantially increasing	37.4	There a shortage of consumables	48.1

According to PM respondents, parameters in the “project practices” (PM practices and project best practices) and global categories were the most frequent in the top 10 parameter lists for positive and negative effects, respectively. “Project practice” parameters, such as the “application of integration management”, “zero accident techniques”, and “cost management practice (reporting system for identification of cost overruns)” comprised three of the top ten parameters for positive effect; global parameters like “uncertainty of global economy”, “low natural gas prices”, and “volatility of oil price” comprised three of the top ten parameters for negative effect on CLP, which is expected in energy driven economies like that of Alberta and Canada (Chanmeka et al. 2012; Chan et al. 2004).

For trade respondents, parameters related to “labour and crew” comprised six of the top ten parameters for positive effect, confirming the findings of El-Gohary and Aziz (2014), where “labourer experience and skill” was ranked first among 30 factors, while parameters in the “material and consumables” and “location property” categories comprised seven of the top ten parameters having negative effect on CLP. According to Thomas et al. (1990) and Dai et al. (2009), these categories are manageable on jobsites and can lead to improvements in CLP if properly considered in planning and day-to-day work.

Industrial Context

The rankings of the top 10 parameters, with evaluation scores, identified by PM and trade survey respondents and based on the positive and negative effects on industrial projects are shown in Table 3.7. Of the 141 parameters included in the PM survey, respondents identified “use of daily job hazard assessment system” as the top parameter having a positive effect. This finding substantiates the results obtained by Liberda et al. (2003) where “safety systems including protective gear requirement” was ranked last out of 53 factors based on negative effect on CLP. The PM survey respondents also identified “presence of many competing projects within the province” as the top parameter having a negative effect. Having many competing projects will create a higher demand for construction workers, and a lack of adequate supply could result in the inability to fulfill required human resource needs for projects, misuse of workers skill, and use of unskilled labourers in place of skilled ones, all negatively influencing CLP (El-Gohary and Aziz 2014).

Table 3.7: Industrial Context: Top 10 Parameters

Rank	PM survey Parameters	ES	Trade survey Parameters	ES
	Positive influence:		**Positive influence:**	
1	Daily job hazard assessment system is in place	100	There is really good cooperation between craftsmen in a crew	100
2	There are adequate and quality work tools	82.3	Craftsmen have shown acceptable learning speed	79.6
3	Accidents and injury are infrequent	68.3	For new craftsmen, the job site orientation program is carried out	71.3
4	Cost management practices: A reporting system at company level is in place for the identification of cost overruns	65.0	Craftsmen's labour union status (unionized or not unionized) and its benefits are important in their day to day performance	67.9
5	The organization has many successful years in industry	63.7	Foreman has the required experience	67.2
6	Material order tracking system is in place	62.0	Adequate lunchrooms are closely located	67.0
7	Planning for start-up is being properly carried out	60.3	There is really good cooperation between the different crews	63.0
8	Daily project briefing and debriefing is properly practiced	55.5	Crew is experienced and has the necessary competence	62.0
9	Delivered materials are of high quality	51.7	The work area is protected from weather effect	60.7
10	Quality management practices: Identifying quality requirements and/or standards and documentation on compliance properly implemented	49.8	Frequency of accidents and personal injury is low	59.4
	Negative influence:		**Negative influence:**	
1	There are many competing projects within the province	100	Work conditions are compromised by excessive noise, dust and fumes	100
2	Global economy still faces uncertainty of facing another slow down	80.0	Materials are not delivered on time to task location	96.1
3	Prices for outputs (project completion costs) are substantially increasing	76.7	Workers cannot access the required power tools to do their jobs	90.8
4	Canada's population is aging	70.1	Workers cannot access the required hand tools to do their jobs	79.1
5	Work locations are confronted with excessive noise, dust, and fumes	68.5	The materials delivered have quality problems	69.6
6	The available labour for construction projects in Alberta is inadequate	60.1	Electrical power gets disconnected during operation	68.9
7	Oil prices (dollar/barrel) are highly volatile	47.7	Correction work due to quality problems of prefabricated products is necessary	66.3
8	Recession of global economy is expected in the near future	47.3	There a shortage of consumables	60.4
9	Drawings and specifications are often not complete and require updates	40.4	There is a shortage of good transportation equipment (cranes, forklifts)	46.0
10	Owners are frequently suspending projects	37.0	Washrooms are not closely located	43.8

Of the 89 parameters included in the trade survey, respondents identified “good cooperation between craftsmen in a crew” as the top parameter having a positive effect on CLP. This outcome is supported by Jergeas (2009), whose work identified “labour relations” as the top target area for labour productivity improvement in industrial projects, and Liberda et al. (2003) who ranked it 12th among 33 factors in the “management” category. The trade survey respondents also identified “work conditions compromised by excessive noise, dust, and fumes” as the top parameter having a negative effect. The rankings of the top 10 key parameters shown in Table 3.7 provide deeper insight into parameters influencing industrial projects.

According to PM respondents, and similar to the building context, parameters in the “project practices” (PM practices and project best practices) categories and the “global” category most frequently appeared in the top 10 parameter lists for positive and negative effects, respectively. “Project practice” parameters, specifically “application of cost management practice (reporting system for identification of cost overruns)”, “material order tracking system”, “proper planning for startup”, and “quality management (identifying quality requirements and/or standards and documentation of how the organization will demonstrate compliance)”, comprised four of the top ten parameters for positive effect; global parameters, specifically “uncertainty of global economy”, “volatility of oil price”, and “expected recession of global economy”, comprised three of the top ten parameters for negative effects on CLP. The result related to the global economy parameters and their influence on CLP is in line with the finding by Chan et al. (2004), which identified “economic environment” as a key factor contributing to the success of construction projects. Similar to the trade respondents in the building context, parameters related to “labour and crew” comprised six of the top ten parameters having positive effects on CLP. Eight of the top ten parameters rated as having negative effects on CLP included those in the “material and consumables” and “equipment and tools” categories, which is in line with the CII study on industrial projects (CII 2006; Dai et al. 2009).

3.3.2.3: Overall Key Parameters Influencing CLP

This research investigated the positive and negative influence of 169 parameters on CLP using PM and trade surveys. The overall most significant parameters influencing CLP either positively or negatively, regardless of project context, are as follows: (1) adequate and quality work tools; (2) aging of Canada's population; (3) job site orientation program for new craftsmen; (4) lack of protection from weather effect; (5) use of daily job hazard assessment system; (6) presence of many competing projects within the province; (7) good cooperation between craftsmen in a crew; and (8) work conditions compromised by excessive noise, dust, and fumes. These top eight parameters were extracted from both industrial and building context results shown in Table 3.6 and Table 3.7, where the first ranked parameters for positive and negative influence are taken from both the PM and trade survey respondent groups. Notably, out of the top eight ranked parameters, five are at the activity level and are related to "labour and crew", "equipment and tools", and "location property". This finding is consistent with the findings of CII (2006), Dai et al. (2009), Jarkas and Bitar (2012), and Eslamdoost and Heravi (2013); however, it is in contrast to the studies by Liberda et al. (2003) and El-Gohary and Aziz (2014), which indicated that "management" level factors were ranked higher than "human/labour" factors.

3.3.2.4: Comparison of Perspectives on Parameters between Project Managers and Trades within the Same Context

As the parameters influencing CLP are multilevel, it is important to determine instances where there is a lack of consensus on parameters between PM and trade personnel so as to formulate effective improvement strategies. While a higher level of agreement on parameters between the two groups will help in implementing improvement strategies, a lack of agreement will demand further investigation into the sources of difference before taking action. Thus, in order to investigate the differences in perspective between PM and trade survey respondent groups, the 37 common parameters shown in Table 3.2 were evaluated for differences in perspective. As the respondents rank parameters based on positive and negative effect on CLP, the comparison was made for each parameter based on positive and negative evaluation scores, and parameters with the greatest positive and negative evaluation score differences between the two perspectives were identified. Then, the impact ratings of the identified parameters from both perspectives were again used for statistical analysis and the results were subsequently used to compare group means using an F-test. The significance of the differences was reviewed using the null hypothesis that there is no difference between the groups, which will be rejected at a significance level (p-value) of 5% (i.e., 95% confidence level). The results of the investigation into the perspectives of the different respondents are shown in Appendix B.2, where the top five parameters having the greatest positive and negative evaluation score differences between the two perspectives have been identified for discussion.

Results at the 95% confidence level indicate that the PM and trade groups perceived some parameters' influences on CLP very differently. Statistically significant differences were reported between different generations of craft workers (Dai and Goodrum 2012), foremen and craft workers (Dai et al. 2007), union and non-union craft workers, and trades (Dai et al. 2009). Eslamdoost and Heravi (2013) also observed similar differences among office and site staff respondents in comparing the rankings of factors influencing CLP. In the building context, according to respondents' rankings of positive effect parameters, the most significant difference in perception between the PM and trade groups was found for the "crew experience and competence" parameter, which the trade group ranked highly for positive effect.

According to respondents' rankings of negative effect parameters, the most significant difference in perception between the PM and trade groups was found for "harshness of weather", which the trade group ranked highly for negative effect. In the industrial context, according to respondents' rankings of positive effect parameters, the most significant difference in perception between the PM and trade groups was found for the "good cooperation between craftsmen in a crew" parameter, which the trade group ranked highly for positive effect. According to respondents' rankings of negative effect parameters, the most significant difference in perception between the PM and trade groups was found for "workers not getting required hand tools to do their jobs", which the trade group ranked highly for negative effect.

Most importantly, all of the top five parameters perceived as having negative effects and having the greatest evaluation score differences between the PM and trade groups in the industrial context were related to "material and consumables" and "equipment and tools"; in all cases only trade respondents rated them highly for negative effect. These findings on perspective differences are important for project productivity management and improvement; the parameters with greater differences in perspectives should be further investigated, as trade workers provide detailed insight into the parameters influencing their daily productivity and their opinions are critical for CLP analysis and improvement (Dai et al. 2007).

3.3.2.5: Comparison of Perspectives between Project Management and Trade Respondents between Contexts

Further analysis was conducted to examine differences in perspective between similar respondents from the two different contexts, building and industrial. Such analysis is often missing in CLP studies, even though it could result in a better understanding of the contextual differences between respondents (Dai and Goodrum 2012). The findings of the analysis can improve our understanding of which parameters are specific to each context. Similar to the analysis shown above, to compare perspectives on parameters within a context, the top five parameters having the greatest differences in perspective between the two contexts (building and industrial) for both the PM and trade perspectives have been identified for discussion and are illustrated in Appendix B.3.

Results at the 95% confidence level indicate that respondents from the two contexts perceived some parameters' influences on CLP very differently. According to the PM survey respondents' rankings of positive effect parameters, the most significant difference in perception between the building and industrial context PM groups was found for the "presence of adequate and quality work tools" parameter, which was ranked highly for positive effect in the building context. According to the PM survey respondents' rankings of negative effect parameters, the most significant difference in perception between contexts was found for "presence of many competing projects within the province", which was ranked highly for negative effect in the industrial context.

According to the trade survey respondents' rankings of positive effect parameters, the most significant difference in perception between the different contexts was found for "protection of work area from weather effect", which was ranked highly for positive effect in the industrial context. According to the trade survey respondents' rankings of negative effect parameters, the most significant difference in in perception between the different contexts was found for "workers not getting required hand and power tools", which was ranked highly for negative effect in the industrial context.

In summary, using a focus group representing different construction companies, the hierarchal list of 169 parameters influencing CLP was first verified. The factor survey respondents from the different construction projects under building and industrial contexts examined and verified the relevance of the 169 hierarchal parameters. The extent to which each parameter exists in the given project setting was evaluated by the experts using the agreement or frequency sections of the factor surveys, and the results showed that the parameters do exist in each of the studied project, but with varying degrees of occurrence.The comparison among PM and trade workers within the same context showed difference in perception; which indicated that data on parameters at activity level could be provided by trade level workers (crew members and foreman) and parameters at project level and higher could be provided by PM workers (superintendent and project managers). Additionally, the comparison of the building and industrial context groups' perceptions of parameters influencing CLP were compared, and significant differences were observed, indicating the context-specific nature of parameters influencing CLP.

3.4: Data-Driven Approach for Identification of Key Input Parameters of CLP

The results of the expert-driven approach indicated that all parameters have to be documented using data sources at different levels of the project. Therefore, in this research input parameter documentation was carried out for all input (factors and practices) parameters together with CLP values. The collected data was then used to identify key parameters for CLP modeling using data analysis methods. Data-driven approaches for identifying key variables or features mainly focus either on feature extraction or feature selection techniques and are commonly used to reduce features of datasets (Guyon and Elisseeff 2003). In modeling studies, the predictive power of developed models is hampered when the number of features increases, commonly referred as the curse of dimensionality. Feature extraction techniques like principal component analysis build derived features from existing ones so as to reduce the dimensionality of the feature space and improve model performances; while feature selection reduces the dimensionality by selecting only a subset of measured features and is recommended when the original units and meaning of features are important and the modeling goal is to identify an influential subset.

In this research, instead of feature extraction, feature selection is used, as feature extraction would create new sets of features, or parameters, in addition to the existing ones, and these new features would be less informative as they would not have the same meaning as the original set of features. Feature selection produces an optimal subset of relevant parameters and is often used in data mining areas such as classification, clustering, association rules, and regression (Saeys et al. 2007). Feature selection generally requires two main processes: feature subset evaluator and search method (Liu and Yu 2005). Feature sets are assessed for their merit for explaining the target or output variable using the feature subset evaluator process, while the search method process facilitates a search of all possible subsets. Filter, wrapper, and hybrid methods are the most commonly used feature subset evaluators (Liu and Yu 2005). Filter methods, the most commonly used feature selection method, evaluates and ranks features without involving any mining algorithm (Saeys et al. 2007). Wrapper methods requires one pre-determined mining algorithm and assess subsets of the features according to their suitability to the mining algorithm and improve the prediction performance; while, hybrid methods perform feature selection by combining the different filter and wrapper evaluation criteria in different search stages (Guyon and Elisseeff 2003).

Classically having features in the range of hundreds will result into a large dimension problem and in cases where the available data instances are relatively limited, the use of filter methods is recommended (Xing et al. 2001). As discussed in Section 2.3.1 and shown in Table 2.1, a total of 169 input parameters influencing CLP are identified; additionally, the data instances collected for the trades under study are limited (refer to Table 2.9). Therefore, in this study filter based feature selection algorithm as shown in Table 3.8 was adopted for identifying the key influencing parameters. The filter algorithm, for a given data set *D* of activity input and output parameter, starts the search from a given subset S_0 and searches through the feature space using a search strategy; while each generated subset *S* is evaluated by an independent measure *M* and compared with the previous best one (Liu and Yu 2005). Accordingly, if the new subset is found to be better, it is regarded as the best subset, and the search is iterated until a predefined stopping criterion δ is met and the resulting best subset S_{best} provides the key input parameters influencing CLP.

Table 3.8: Generalized Filter Algorithm (Liu and Yu 2005)

input:	$D\ (F_0, F_1, \ldots, F_{k-1})$ // a training data set with k features, where k = 169 for input parameters
	S_0 // a subset from which to start the search
	δ // a stopping criterion
output:	S_{best}

```
begin
    initialize: S_best = S_0;
    γ_best = eval(S_0, D, M); // evaluate S_0 by an independent measure M
    do begin
        S = generate (D); // generate a subset for evaluation
        γ = eval(S, D, M); // evaluate the current subset S by M
        if (γ is better than γ_best)
            γ_best = γ;
            S_best = S;
    end until (δ is reached);
    return S_best;
end;
```

Correlation-based feature selection (CFS) algorithm is a simple and powerful filter method that evaluates the relevance of features using Pearson correlation coefficient (Hall 1998). CFS has been proven to perform very well in experiment with small datasets (Hall 1998). In this research the use of the CFS algorithm was found to be suitable, as it has the ability to deal with a high dimension of the parameters influencing CLP and small number of data instances; while preserving the original representation of the parameters and providing better understanding of the underlying process that generated the data (Guyon and Elisseeff 2003). CFS ranks feature subsets based on heuristic evaluation function shown Eq. (3.19), where M_s is the heuristic “merit” for the feature subset *S* containing *k* features, $\overline{r_{cf}}$ is the mean feature to output (i.e., class or CLP) correlation $(f \in S)$, and $\overline{r_{ff}}$ is the average feature to feature inter-correlation (Hall 1998):

$$M_s = \frac{k\ \overline{r_{cf}}}{\sqrt{k + k\ (k-1)\ \overline{r_{ff}}}} \tag{3.19}$$

According to Hall (1998), the numerator in Eq. 3.19 can be considered as providing an indication of how predictive of the output or class a set of features are, while the denominator indicates how much redundancy there is among the features. The correlation between features and output or among features is computed using Pearson’s correlation coefficient as shown in Eq. 3.20—Eq. 3.25; where $x_1, x_2, \ldots, x_j$ represents the subset containing P features (input parameters) and y represents the output

(CLP), $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_j$, and $\bar{y}$ are the mean values, $cov(x_j, x_{j+1})$ is the covariance of any two features and $cov(x_j, y)$ is the covariance of any one of the features and output, N are the number of data instances, and $\sigma_{x_1}, \sigma_{x_2}, \dots, \sigma_{x_j}$ and σ_y are standard deviations:

$$cov(x_j, x_{j+1}) = \sum_{i=1}^{N} \frac{(x_{ji} - \bar{x}_j)(x_{j+1i} - \overline{x_{j+1}})}{N} \tag{3.20}$$

$$cov(x_j, y) = \sum_{i=1}^{N} \frac{(x_{ij} - \bar{x}_j)(y_i - \bar{y})}{N} \tag{3.21}$$

$$r_{x_j x_{j+1}} = \frac{cov(x_j, x_{j+1})}{\sigma_{x_j}\, \sigma_{x_{j+1}}} \tag{3.22}$$

$$r_{x_j y} = \frac{cov(x_j, y)}{\sigma_{x_j}\, \sigma_y} \tag{3.23}$$

$$\overline{r_{ff}} = \frac{\sum_{j=1}^{P} r_{x_j x_{j+1}}}{P - 1} \tag{3.24}$$

$$\overline{r_{cf}} = \frac{\sum_{j=1}^{P} r_{x_j y}}{P} \tag{3.25}$$

Three search strategies are recognised: complete, sequential, and random. In complete search strategy an exhaustive search is carried out and optimal results are guaranteed; however, the search space, due to the large dimension in CLP modeling, will be quite large to handle with limited data instances. Additionally, the use of simple heuristic based sequential search strategies (e.g. BestFirst) have been found to be as efficient as random search strategies based on genetic algorithm and simulation annealing (Rodrigues et al. 2015). Thus, in this research BestFirst sequential search strategy has been adopted. BestFirst searches the feature space using a greedy hill climbing approach augmented with a backtracking facility (Xu et al. 1988), and the search strategy starts either from empty set of features and searches forward, or starts with the full set of features and searches backward, or starts at any point and searches in both directions (by considering all possible single feature additions and deletions at a given point). The search stops when subsequent addition (or deletion) of any feature does not produce a better subset (Liu and Yu 2005).

3.4.1: Implementation of Data-driven Approach

A number of open source and commercial programs are available for implementing the data-driven approach for selecting key parameters using feature selection. In this research, the most relevant features were selected using the Waikato Environment of Knowledge Analysis (WEKA) tool. Accordingly, WEKA's correlation-based feature selection (CfsSubsetEval) algorithm was employed as the feature subset evaluator with the BestFirst algorithm for the search method, as together, they provide the most efficient approach to correlation-based feature selection (Rodrigues et al. 2015).

3.4.2: Key Parameters Influencing Concreting Activity

The construction industry is constantly searching for ways to improve labour productivity, as labour productivity significantly impacts the project costs and profitability of construction companies (Fulford and Standing 2014). However, before they can propose and implement improvement strategies, industry representatives need an activity-level construction labour productivity model that enables them to fully understand which parameters (factors and practices) cause productivity to change and by how much (Thomas et al. 1990). The data-driven approach relies on the field data for construction labour productivity and the influencing parameters; therefore, the approach is applicable to identification of parameters at activity level, where labour productivity values are studied, and provides the much needed key parameters for CLP model development.

In this section, the identification of the key parameters influencing the concreting (concrete placement) activity is presented. The analysis focused on concreting data collected from six building projects. Data instances from the six structural elements were combined, which, compared to the other activities studied (also shown in Table 2.9), produced the largest data set with a total of 92 data instances. Using the WEKA tool and CfsSubsetEval algorithm together with the BestFirst search algorithm, 14 key parameters, shown in Table 3.9, were identified from a total of 105 input parameters, which were made up of recorded factors and practices. Key input parameters (factors and practices), or features having high correlations with CLP but low inter-correlations, were identified. The full detail of the correlation results among key input parameters and CLP is shown in Appendix B.4.

Table 3.9: Key Input Parameters Influencing Concreting Activity: Pearson Correlation with CLP

ID	Parameter (ID)	Scale of measure	Correlation coefficient
x97	Owner staff on site	Integer (total number of owner staff on site)	-0.343 [a] (0.001)
x143	Availability of communication devices	Real number (ratio number communication radios: total number of crews)	-0.289 [a] (0.005)
x13	Craftsperson trust in foreman	1–5 predetermined rating	0.273 [a] (0.008)
x11	Craftsperson motivation	1–5 predetermined rating	-0.261 [a] (0.012)
x96	Project work times	Real number (ratio total worked hours per day)	0.227 [a] (0.030)
x46	Structural element type	Categorical: columns (1), footings (2), grade beams (3), pile caps (4), slabs (5), walls (6)	0.227 [a] (0.30)
x126	Labour productivity measurement practice	1–5 predetermined rating	0.218 [a] (0.037)
x101	Approval of shop drawings and sample materials	Real number (average time taken to approve, days)	-0.175 (0.095)
x12	Craftsperson fatigue	Real number (ratio total weekly worked hours: regular weekly work hours)	-0.143 (0.173)
x58	Change of foremen	Integer (number of turnovers per month)	-0.135 (0.199)
x15	Level of absenteeism	Real number (ratio absent crew member: total crew size)	0.099 (0.346)
x37	Availability of tools	Real number (average waiting time, minutes)	0.091 (0.387)
x45	Concrete placement technique	Categorical: pump (1), crane and bucket (2), direct chute (3)	-0.078 (0.461)
x23	Job security	Real number (average length of unemployment period, months)	-0.073 (0.487)

Note: Values in parentheses indicate the significant value for a two-tailed correlation test
[a] Denotes a statistically significant correlation at a significance level of 0.05.

The data-driven results indicate that the owner staff on site, availability of communication devices, craftsperson trust in foreman, craftsperson motivation, project work times, structural element type, and labour productivity measurement practice parameters significantly correlated to the actual productivity and its variability. Using the signs of the correlation coefficient the positive and negative impact of the parameters can also be evaluated. The results showed some similarity with the top 10 parameters for building context, shown in Table 3.6, where adequacy and quality of work tools (PM survey), taking efforts to reduce foremen turnover (PM survey), drawings and specification are often not complete and require updates (PM survey), and craftsmen trust in the skills and judgment of their supervisor (Trade survey) were identified as a key parameters.

However, the key parameters established using data-driven approach are specific to an activity, while the expert-driven approach is applied at project level using opinions of project participants at different level. Thus, the key parameters identified using the two approaches are rather complementary and provide a better oversight of the parameters influencing CLP. Additionally, the key parameter lists for building and industrial contexts, based on the expert-driven approach, have shown differences; which indicated that key parameters influencing CLP are context specific. This finding confirms with findings of previous studies which recognized the context specificity of key parameters, and documented project features like building type, project location, contract type, union type, etc. to contextualize the studied parameters (Christian and Hachey 1995; Thomas and Raynar 1997).

3.5: Chapter Summary

The construction industry, being project-based in execution, is exposed to a complex and unpredictable setting, or context, which influences the efficiency of construction labour productivity (CLP) either directly or indirectly. This being the case, one major problem in CLP studies is identifying key influencing parameters that are relevant for the context under study. Such an effort requires careful consideration of the parameters influencing CLP and their multilevel, complex, and context-dependent nature. Accordingly, in this thesis a comprehensive list of 169 parameters (factors and practices) influencing CLP was established based on a detailed review of past studies addressing various contexts, and was verified using experts from various companies carrying out several building and industrial projects. To examine and elucidate the interlinked and context-dependent nature of the parameters influencing CLP, this study evaluated the positive as well as the negative influences of the parameters based on survey responses from PM and trade level participants from a number of ongoing construction projects categorized under two contexts: building and industrial. Unlike previous studies on CLP, the survey questions in this study focused on specific project contexts and examined differing perspectives of project managers and trades on a given project. The results of the survey analysis established ranked lists of parameters for the building and industrial project contexts; additionally, the differences in perspectives between the PM and trade level respondents were examined for each given context as well as between the two contexts. Using the results of the expert-driven, a data-driven feature selection approach was applied to collected field data; and the key parameters influencing concreting (concrete placement) activity were identified.

The results of this chapter demonstrate the complex and interlinked nature of the parameters influencing CLP, and the need and importance of identifying the key parameters based on positive and negative influence, and using both expert and data-driven approaches. They also confirm that PM and trade workers within the same context perceive the influence of some parameters on CLP very differently, and that the key parameters influencing CLP are context-dependent. These findings indicate that in order to formulate effective CLP analysis models and, accordingly, CLP improvement strategies, studies must include different levels of project personnel and clearly establish the context from which the key parameters were identified. Accordingly, the context-specific nature of the key parameters will lead to the development of CLP models which are specific to the studied context. In the next chapter, the effect and role of the other influencing process parameters or work sampling proportions on CLP is formulated, and an approach for integrating the key input and process variables using a system based CLP modeling approach is proposed, developed, and tested.

3.6: References

Chan, A., Scott, D., and Chan, P. (2004). “Factors affecting the success of a construction project.” *J. Constr. Eng. Manage.*, 130(1), 153–155.

Chanmeka, A., Thomas, S., Caldas, C., and Mulva, S. (2012). “Assessing key factors impacting the performance and productivity of oil and gas projects in Alberta.” *Can. J. Civ. Eng.*, 39(3), 259–270.

Christian, J., and Hachey, D. (1995). "Effects of delay times on production rates in construction." *J. Constr. Eng. Manage.*, 121(1), 20–26.

CII (Construction Industry Institute). (2006). *Work force view of construction labor productivity*, Univ. of Texas at Austin, Austin, Texas, US.

CSC (Construction Sector Council). (2013). *Construction looking forward, national summary, 2013–2021 key highlights*, Construction Sector Council, Ottawa, Ontario, Canada.

Dai, J., and Goodrum, P. M. (2012). “Generational differences on craft workers' perceptions of the factors affecting labour productivity.” *Can. J. Civ. Eng.*, 39(9), 1018–1026.

Dai, J., Goodrum, P. M., and Maloney, W. F. (2007). “Analysis of craft workers' and foremen's perceptions of the factors affecting construction labour productivity.” *Constr. Manage. Econ.,* 25(11): 1139–1152.

Dai, J., Goodrum, P. M., and Maloney, W. F. (2009). “Construction craft workers’ perceptions of the factors affecting their productivity.” *J. Constr. Eng. Manage.*, 135(3), 217–226.

Durdyev, S., and Mbachu, J. (2011). "On-site labour productivity of New Zealand construction industry: Key constraints and improvement measures." *Aust. J. Constr. Econ. Building*, 11(3), 18–33.

El-Gohary, K.M., and Aziz, R.F. (2014). “Factors influencing construction labor productivity in Egypt.” *J. Manage. Eng.*, 30(1), 1–9.

Engwall, M. (2003). "No project is an island: linking projects to history and context." *Res. Pol.,* 32(5), 789–808.

Eslamdoost, E., and Heravi, G. (2013). “Identification and evaluation of effective factors on labor productivity in power plant construction projects.” *Proc. Construction Specialty Conference*, CSCE, Montreal, Quebec, Canada, 10 pp.

Fulford, R., and Standing, C. (2014). “Construction industry productivity and the potential for collaborative practice.” *Int. J. Project Manage.*, 32(2), 315–326.

Gerek, İ. H., Erdis, E., Mistikoglu, G., and Usmen, M. (2014). “Modelling masonry crew productivity using two artificial neural network techniques.” *J. Civ. Eng. Manage.*, (ahead-of-print), 8pp.

Guyon, I., and Elisseeff, A. (2003). “An introduction of variable and feature selection.” *J. Mach. Learn. Res., 3*, 1157–1182.

Hall, M. A. (1998). “Correlation-based feature selection for machine learning.” PhD thesis, Univ. of Waikato, Hamilton, New Zealand.

Jarkas, A., and Bitar, C. (2012). “Factors affecting construction labor productivity in Kuwait.” *J. Constr. Eng. Manage*., 138(7), 811–820.

Jergeas, G. (2009). “Improving construction productivity on Alberta oil and gas capital projects.” Alberta Finance and Enterprise, Alberta, Canada.

Knight, K., and Fayek, A. Robinson (2000). “A preliminary study of the factors affecting the cost escalation of construction projects.” *Can. J. Civ. Eng.*, 27(1), 73–83.

Liberda, M., Ruwanpura, J., and Jergeas, G. (2003). "Construction productivity improvement: A study of human, management and external issues." *Proc., Construction Research Congress*, ASCE, Honolulu, Hawaii, 1–8.

Lim, E. C., and Alum, J. (1995). "Construction productivity: issues encountered by contractors in Singapore." *Inter. J. Proj. Manage*., 13(1), 51–58.

Liu, H., and Yu, L. (2005). “Toward integrating feature selection algorithms for classification and clustering.” *IEEE T.. Knowl. Data Eng.,* 17(4), 491–502.

Moselhi, O., and Khan, Z. (2012). “Significance ranking of parameters impacting construction labor productivity.” *J. Cons. Innov.: Info. Proc. Manage.*, 12(3), 272–296.

Oduba, A. O. (2002). “Predicting industrial construction productivity using fuzzy expert systems.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Panas, A., and Pantouvakis, J. P. (2010). “Evaluating research methodology in construction productivity studies.” *Built Hum. Env. Rev.,* 3(1), 63–85.

Rodrigues, D., Prata, D., and Silva, M. (2015). “Exploring social data to understand child labor.” *Int. J. Social Sci. Humanity*, 5, 29–33.

Saeys, Y., Inza, I., and Larranaga, P. (2007). “A review of feature selection techniques in bioinformatics.” *Bioinformatics,* 23(19), 2507–2517.

Song, L., and AbouRizk, S. M. (2008). "Measuring and modeling labor productivity using historical data." *J. Constr. Eng. Manage*.,134(10), 786–794.

Sonmez, R., and Rowings, J. E. (1998). "Construction labor productivity modeling with neural networks." *J. Constr. Eng. Manage*.,124(6), 498–504.

Stewart, T. J., and Frye, A. W. (2004). “Investigating the use of negatively-phrased survey items in medical education settings: Common wisdom or common mistake?”, *Acade. Medicine*, 79(10), S1–S3.

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). "Modeling construction labor productivity." *J. Constr. Eng. Manage*., 116(4), 705–726.

Thomas, H. R., and Raynar, K. A. (1997). "Scheduled overtime and labor productivity: quantitative analysis." *J. Constr. Eng. Manage*., 123(2), 181–188.

Tsehayae, A. A., and Fayek, A. Robinson. (2012). “A research framework for work sampling and its application in developing comparative direct and support activity proportions for different trades,” *Proc., Construction Research Congress*, ASCE, West Lafayett, Indiana, US, 1420–1429.

Tsehayae, A. A., and Fayek, A. Robinson. (2014a). “Data-driven approaches to discovering knowledge gaps related to factors affecting construction labour productivity,” *Proc., Construction Research Congress*, ASCE, Atlanta, Georgia, US, 837–891.

Tsehayae, A. A., and Fayek, A. Robinson. (2014b). "Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects." *Can. J. Civ. Eng.,* 41(10), 878–891.

Tsehayae, A. A., and Fayek, A. R. (2014c). *Factor surveys: Interview questionnaires, participation letter, and consent form*. Technical report submitted to Suncor Energy, NSERC IRC SCMD: Productivity/Suncor/ Factor Surveys Report/2014–TD01, Edmonton, Alberta, Canada, October 10, 2014, 40 pp.

Xing, E. P., Jordan, M. I., and Karp, R. M. (2001). “Feature selection for high-dimensional genomic microarray data.” *In ICML*, 1, 601–608.

Xu, L., Yan, P., and Chang, T. (1988). “Best first strategy for feature selection.” *Proc. International Conference on Pattern Recognition*, IEEE, 706–708.

Chapter 4: System Based CLP Modeling[4]

4.1: Introduction

The construction industry is constantly searching for ways to improve labour productivity, as labour productivity significantly impacts the project costs and profitability of construction companies (Fulford and Standing 2014). However, before they can propose and implement improvement strategies, industry representatives need an activity-level construction labour productivity model that enables them to fully understand which parameters (factors and practices) cause productivity to change and by how much (Thomas et al. 1990). Such models also play a key role in construction project estimating, scheduling, and planning decisions (Yi and Chan 2014).

Construction labour productivity (CLP) deals with the efficiency of labour resources in the process of converting input resources, like materials, into the outputs of labour-intensive construction project activities. In this study, CLP, referred to as the output variable (*O*), is defined as the ratio of units of output—in terms of installed quantity—to units of input—in terms of total labour work-hours. Different parameters, made up of various factors and practices (e.g., crew size, safety practice, etc.) are known to affect the conversion process. Of these parameters, this study considers those that critically influence CLP as input variables (*I*) in order to further examine their effects on CLP.

Understanding how time is used during the input-to-output conversion process is also vital to modeling CLP; work-study methods are commonly employed for this purpose. Work sampling, a method used to determine the amount of time workers spend performing direct (productive), supportive, and delay (non-productive) work, is the most widely used work-study method (Josephson and Björkman 2013). Work sampling proportions summarize the actual utilization of labour work hours and are represented as process variables (*P*) in this study; they provide an in-depth examination of what happens during the conversion process as shown in Fig. 1.1.

[4] Parts of this chapter have been submitted for publication in Journal of Construction Innovation: Information, Process, Management, JCI, 36 manuscript pages, submitted July 28, 2015; and have been published in the Proceedings, NAFIPS Annual Meeting 2012, Berkeley, California, US, August 6-8, pp. 1-6.

In order to improve CLP, appropriate analysis and modeling is required so as to clearly illustrate how input variables affect conversion process efficiency. Such an analysis must establish the relationships between the three system model variables—*Input*, *Process*, and *Output* (Fig. 1.1)—so as to examine the cause and effect of the input and process variables on CLP. The main objective of this chapter is to present a novel system-based approach for modeling the output variable (*O*) (i.e., CLP) using input (*I*) and process (*P*) variables. This chapter begins with a literature review of past approaches used for modeling CLP. Next, it discusses the research methodology used to collect data related to the system model variables. Subsequently, using extensive field data collected for eight activities, it examines the validity of activity models by testing the relationship between the output (CLP) and process variables so as to verify the usefulness of relying on work sampling proportions like direct work (also called "tool time") in the prediction of CLP. Finally, this chapter presents a novel system model of CLP using input variables in conjunction with process variables, tests the approach using field data, and then summarizes conclusions.

4.2: Background

Because of its significance to project performance, CLP has been well studied. As discussed in detail in Section 2.2, numerous modeling approaches have been developed and tested. Overall, these tested approaches can be categorized as either factor or activity models. Factor models related the different input variables—made up of key influencing parameters (factors and practices) to labour productivity. Activity models mainly related the process variables, in terms of work sampling proportions, to labour productivity. Most of the past CLP studies have quantified the impact of different parameters on CLP using factor models. The relationship between the assorted input variables and CLP, and the degree of their impact on it, has most frequently been modeled using regression analysis. Thomas et al. (1994) developed a factor model for masonry projects to forecast the labour productivity of masons. More recent CLP studies focus on the use of neural networks (NN). Gerek et al. (2015) developed NN models to predict the productivity of masonry crews. Because it is a complex problem with limited data availability and deals with a large number of subjective and linguistic variables, CLP is an exceptional target for fuzzy set modeling. However, there are few applications of fuzzy sets in the CLP modeling field (Yi and Chan 2014). One such application is the fuzzy inference based labour productivity model Oduba (2002) developed for industrial pipe rigging and welding activities.

Activity models relate labour utilization measures like work sampling proportions (i.e., process variables) to CLP. Work sampling (WS), a widely used work-study method in the construction industry, uses random observation to investigate how a workforce uses its work time. Craftsman questionnaires and foreman delay surveys have been used to both supplement and complement WS (CII 2010). Such surveys are useful for identifying the causes of delays (e.g., rework, equipment breakdown, waiting for instruction, etc.) and quantifying the resulting lost labour work-hours. WS establishes the percentage of work time spent on WS categories like direct work; which represents the proportion of work time spent exerting physical effort directed toward the completion of an activity, has been used as a surrogate measure of CLP (Thomas et al. 1990). The activity model is based on WS and is readily applicable to labour-intensive activities. A valid activity model is required to show that direct work times and outputs are related in some predictable fashion (Thomas 1991). There are two opposing views of the validity of activity models. For the most part, the literature argues that WS can be used to predict productivity (Shahtaheri et al. 2015). Handa and Abdalla (1989) stated that activity models with 10 WS categories could be used to indicate actual site productivity; however the developed models were statistically insignificant. Silva and Ruwanpura (2006) developed an activity model using direct work proportion to predict concreting operations' productivity. Shahtaheri et al. (2015) assumed labour performance can be improved by increasing the direct work rate and the developed baselines for direct work proportions without testing and verifying the relationship between direct work and CLP. Conversely, Thomas (1991) stated that direct work is not related to productivity; using data from seven databases containing over 158 WS studies, mainly from nuclear power plant projects, he concluded that previous studies lacked validity. Josephson and Björkman (2013) also concluded that WS studies provide little value in measuring productivity.

Most of the factor models were not able to deal with subjective variables in a comprehensive manner, studied limited number of influencing parameters, and focused on non-interpretable neural networks for developing CLP models. Furthermore, activity models did not show statistically significant results, and were limited by their inability to model the effect of the parameters influencing CLP and, due to a lack of standardization, depended on assumptions regarding WS category definitions (Josephson and Björkman 2013). So far, no previous studies have succeeded in developing an integrated system approach investigating the overall relationship between both input and process variables and CLP. As a result of these limitations, the above models have been unable to provide useable solutions for the highly complex, context-dependent, and non-linear modeling problem of CLP.

4.3: Research Methodology for Developing the System (I–P–O) Variables

This chapter proposes a novel approach to overcome the limitations and challenges of CLP analysis and modeling through the use of a system-based model. Accordingly, it aims to accomplish the selection of key parameters that, together with work sampling proportions, help explain the variability in CLP. In this subsection the research methodology, also described in Section 2.3, used to collect data on the system model variables—namely, input, process, and output variables, is briefly summarized.

The vital and starting point of any CLP analysis and modeling study involves the identification, quantification, and documentation of parameters influencing CLP and then the establishment of key parameters. These key parameters make up the *input variables* for models that can then be used to develop improvement opportunities. Using the identified 169 parameters (factors and practices) influencing CLP (shown in Table 2.1), and the measurement scheme described in Section 2.3.1, the input parameters were documented. Additionally, one major objective of this research is to test the validity of using *process variables*, based on the key work sampling proportions, for explaining variance in CLP. For the first time in CLP research, this study properly investigates the relationship between process variables and input variables. This chapter presents insights into how input variables influence process efficiency and, consequently, CLP. This research focused on crew-level WS by studying as a whole the specific crew performing a given activity. The "modified crew" method of WS is adopted, by which WS observation focuses on the study of representative crews performing selected activities (CII 2010). In this study, process parameters are collected using the CII *Guide to activity analysis* (2010); accordingly, the CII's seven standard categories—direct work, preparatory work, tools and equipment, material handling, waiting, travel, and personal—are used in WS observations. In order to achieve statistically significant process parameters and taking into account the seven WS categories and a confidence level of 95% with a margin of error of ±5%, a total of 510 observations were targeted per hour of the study (Gouett et al. 2011). Activities were studied in three data collection cycles, where each cycle extends over a month so as to attain the required number of observations. After all the hourly observation periods of a given day have been completed, the total head counts for each WS category are tabulated. The proportion of time spent on each WS category is then calculated for each crew by taking the ratio of their respective number of observations to the total number of observations.

Determination of the *output variable*—CLP—was based on the ratio of output (installed quantity) to input (total work-hours), as shown in Eq.(1.1). The labour productivity data includes details about the size of the crew performing the task, the total man-hours, and the installed quantity. The actual size of a crew on a given day was determined according to the number of workers present, which was verified by comparing it with the crew size observed during WS. The total man-hours were based on the actual craft work hours spent by a crew, and were computed based on the sum of the recorded activity work time for each of the crew members. The lost hours values determined using the foreman delay surveys were deducted from the crew total work-hours used to calculate productivity (refer to Eq. (1.1)) if verified by crew members via the craftsman questionnaires. The installed quantity was collected using a measurement of units, as in the projects in this study, the relevant activities could be completed in less than a shift and so counting the activity units completed was easily and accurately done.

4.4: TESTING THE VALIDITY OF ACTIVITY MODELS (P–O RELATIONSHIP)

This section examines the validity of activity models by testing the relationship between the output (CLP) and process parameters, so as to verify the usefulness of relying on work sampling proportions in the prediction of CLP. First, a correlation analysis is carried out between the process parameters and CLP data. Next, validity of activity models based on the relationship between direct work proportions or tool time and CLP is tested. Finally, validity of activity models based on the relationship between the process parameters (seven WS category proportions) and CLP is tested. The analysis was carried out using the IBM SPSS 22® statistical package.

4.4.1: Description of Data

As discussed in Section 2.4, in order to gather adequate data for detailed analysis of CLP, extensive data for input, process, and output parameters were collected from 11 projects across Alberta, Canada. Data collection took place between June 2012 and November 2014 in collaboration with seven partnering companies. The activities were studied in three data collection cycles, where each cycle extended over a month-long period and encompassed different weather seasons. The detailed data-collection protocol was adopted to facilitate data collection by several different collectors and to ensure the validity of the data collected from a number of projects.

As the minimum suggested number of data instances for any regression analysis is 30 (Green 1991), activities that had close to 30 data instances were selected from Table 2.9 for testing the validity of the activity model based on the relationship between process and output variables (i.e., the *P–O* relationship). Thus, a total of eight activities—three from concreting (column, slab, and wall concrete placements), three from electrical (box installation, piping, and wire pulling), and two from shutdown (overlays and shield installation)—were further investigated.

The project type, activities studied, description of the activities, and the number of total data instances used for system model analysis is shown in Table 4.1. A total of 92, 148, and 106 data instances were used from concreting, electrical, and shutdown activity categories, respectively. For each data collection instance, WS observations were made for the crew under study and parameters (factors and practices), total work-hours, and installed quantities were documented. CLP data were computed for each of the eight activities; the mean and standard deviation values are shown in Table 4.2, where higher values are desired. The mean and standard deviation values of the proportions of the seven CII WS categories are also shown in Table 4.2. The direct work proportions observed in the 11 studied projects are in line with the North American industry trend, where direct work proportions are reported to fall between 40 to 60% (CII 2010).

Table 4.1: Profile of Activities Studied for Activity and System Based CLP Modeling

Activity category	Activity	Activity description	Total data instances
Concreting (6)	Columns [a]	Concrete placement for columns	21
	Footings	Concrete placement for footings	5
	Grade beams	Concrete placement for grade beams	6
	Pile caps	Concrete placement for pile caps	2
	Slabs [a]	Concrete placement for slabs	28
	Walls [a]	Concrete placement for walls	30
Electrical (3)	Box installation [a]	Installation of pull and outlet boxes	48
	Piping [a]	Installation of rigid galvanized steel conduit	57
	Wire pulling [a]	Pulling wire	43
Boilermaker (3)	Overlays [a]	Welding additional metal layer on tubes to reinforce their thickness	38
	Shields installation [a]	Replacing shields over tubes by removing old ones and welding new shields over tubes	68

Note: Values in parentheses indicate the total number of studied projects and [a] Denotes the activities used to test the *P–O* relationship

To begin with, using boxplots, outlier data instances were identified, and 4, 5, 6, 13, 7, 11, 9, and 19 outlying data instances were removed from column concreting, slab concreting, wall concreting, box installation, piping, wire pulling, overlays, and shield installation activity data sets, respectively.

Then, a correlation analysis on the seven work sampling activity categories and CLP was carried out. Significance probability was tested using a two-sided t-test at a p-value of 0.05. The results, shown in Table 4.3, reveal that direct work proportion is not significantly correlated with CLP for all eight activities. In all cases, the R-values are quite low. These results are consistent with the observed insignificant correlation coefficient of −0.250 (p = 0.153) in Handa and Abdalla's framing crew study (1989); however, they contradict the 0.90 correlation coefficient reported by a concreting study that used a group-timing technique and only two categories—working or idle (Silva and Ruwanpura 2006). This result dispels the assumption commonly witnessed in CLP research that direct work proportions are highly correlated to construction labour productivity.

Table 4.2: Mean and Standard Deviation of Work Sampling Proportions (%) and CLP Values

Activity	Total data instances	Unit	Direct work	Preparatory work	Tools and equipment	Material handling	Waiting	Travel	Personal	CLP (units/mhr)
Column concreting	21	m^3	37.76 (15.34)	15.24 (10.88)	6.57 (8.94)	9.24 (11.41)	27.48 (16.06)	3.86 (3.88)	0.00 (0.00)	2.61 (1.18)
Slab concreting	28	m^3	46.57 (13.96)	6.79 (6.30)	8.57 (6.69)	1.21 (2.38)	29.32 (12.55)	1.79 (2.17)	5.79 (6.16)	2.27 (0.90)
Wall concreting	30	m^3	49.53 (15.21)	9.20 (7.47)	11.87 (13.14)	3.30 (5.17)	23.77 (10.86)	1.77 (2.50)	0.53 (1.59)	4.97 (3.60)
Box installation	48	ea.	59.87 (13.99)	9.25 (7.98)	3.63 (3.94)	5.15 (5.40)	0.46 (1.49)	7.00 (6.13)	14.83 (9.67)	2.94 (4.64)
Piping	57	m	51.86 (13.34)	12.84 (9.27)	8.61 (5.90)	6.82 (6.14)	1.14 (2.29)	9.16 (4.60)	8.81 (7.89)	3.77 (2.21)
Wire pulling	43	m	53.60 (14.52)	12.30 (11.22)	2.63 (3.81)	5.47 (7.98)	1.09 (3.21)	4.35 (4.86)	21.14 (15.98)	10.93 (11.61)
Overlays	38	inches	62.68 (12.43)	17.05 (11.71)	5.32 (5.63)	0.32 (0.84)	5.74 (5.23)	3.82 (2.50)	5.34 (5.03)	0.61 (1.01)
Shield installation	68	ea.	57.65 (14.29)	19.47 (10.96)	4.82 (4.50)	2.18 (2.62)	9.44 (10.71)	4.04 (3.94)	2.34 (3.44)	0.48 (0.22)

Note: Values in parentheses represent standard deviation.

Table 4.3: Pearson Correlation and Regression Analysis: Work Sampling Proportion with CLP

Activity	Correlation coefficient of independent variables							Model summary		ANOVA	
	Direct work	**Preparatory work**	**Tools and equipment**	**Material handling**	**Waiting**	**Travel**	**Person al**	**R^2**	**$R^2_{adjusted}$**	**F**	**Sig.**
Column concreting N=17	0.454 (0.067)	−0.398 (0.114)	−0.118 (0.651)	−0.109 (0.677)	−0.035 (0.0.895	0.043 (0.871)	c (c)	No significant relationship			
Slab concreting N=23	0.104 (0.652)	0.278 (0.222)	-0.169 (0.464)	0.398 (0.074)	−0.057 (0.806)	0.093 (0.689)	−0.116 (0.618)	No significant relationship			
Wall concreting N=24	0.186 (0.384)	0.001 (0.997)	0.003 (0.989)	−0.465 [a, b] (0.022)	−0.041 (0.850)	0.024 (0.912)	c (c)	0.216	0.180	6.058	0.022
Box installation N=34	0.217 (0.219)	−0.343 [a, c] (0.047)	−0.095 (0.592)	0.165 (0.351)	0.149 (0.399)	−0.158 (0.373)	0.018 (0.918)	0.122	0.095	4.581	0.040
Piping N=50	0.090 (0.535)	−0.330 [a, d] (0.019)	0.208 (0.147)	−0.119 (0.412)	−0.201 (0.161)	−0.025 (0.864)	0.030 (0.836)	0.129	0.111	6.973	0.011
Wire pulling N=32	−0.027 (0.884)	−0.027 (0.882)	−0.098 (0.594)	−0.125 (0.497)	−0.214 (0.240)	−0.076 (0.680)	0.089 (0.628)	No significant relationship			
Overlays N=29	0.317 (0.094)	−0.209 (0.276)	-0.054 (0.779)	-0.272 (0.153)	0.114 (0.556)	-0.334 (0.076)	−0.251 (0.189)	No significant relationship			
Shields installation N=49	0.190 (0.192)	0.005 (0.971)	-0.270 (0.060)	0.077 (0.600)	-0.038 (0.796)	-0.087 (0.551)	-0.123 (0.399)	No significant relationship			

Note: Values in parentheses indicate the significant value for a two-tailed correlation test and *c* indicates that the value cannot be computed as the data instance values are constant.

[a] Denotes a statistically significant correlation at a significance level of 0.05.

[b] The test of coefficients gave a t-test value of 11.267 (p = 0.000) for constant and −2.461 (p = 0.022) for material handling variable.

[c] The test of coefficients gave a t-test value of 8.749 (p = 0.000) for constant and −2.140 (p = 0.040) for preparatory variable.

[d] The test of coefficients gave a t-test value of 9.577 (p = 0.000) for constant and −2.641 (p = 0.011) for preparatory variable.

Of the other six WS activity categories, preparatory work significantly correlates with CLP for the box installation and piping activities with correlation coefficients of −0.343 and −0.330, respectively, indicating a negative effect, even though preparatory work is expected to positively contribute to CLP. Material handling WS category also showed a significant correlation with CLP for wall concreting activity. For all eight activities, the other WS activity categories—namely, tools and equipment, waiting, travel, and personal—do not significantly correlate with CLP.

4.4.2: Direct Work Proportion versus CLP

The fundamental assumption of activity models—that CLP improves if more time is spent on direct work activities—was tested using the data collected for the eight activities shown in Table 4.1 and 4.2. To begin with, scatter plots that considered CLP as the dependent variable and direct work category proportions as the independent variable were prepared. Linear and nonlinear regression lines were fitted to each activity's data. The following nonlinear regression models were tested: logarithmic, inverse, quadratic, compound, power, S, growth, exponential, and logistic. The best nonlinear regression lines showing an improvement over linear regression lines in terms of R^2-values were selected. The scatter plots for column concreting and wire pulling activities are shown in Fig. 4.1 and Fig. 4.2.

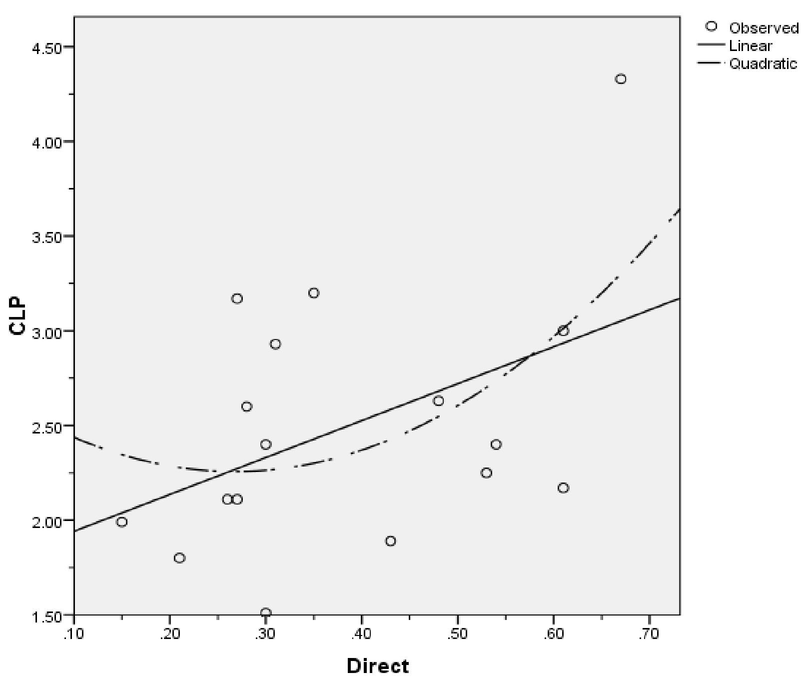

Figure 4.1: CLP as Function of Direct Work Proportion for Column Concreting Activity

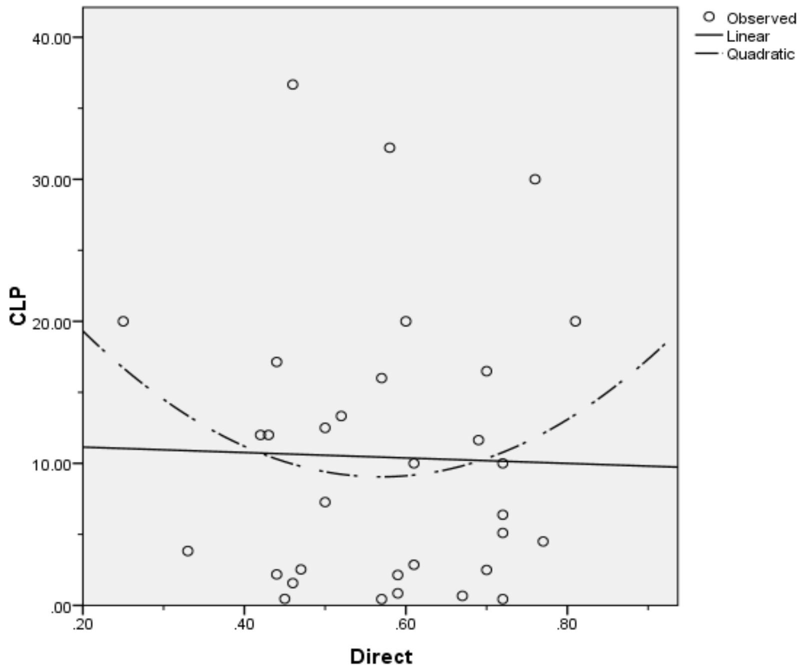

Figure 4.2: CLP as Function of Direct Work Proportion for Wire Pulling Activity

The example scatter plots shown in Fig. 4.1 and Fig 4.2 indicate that an increase in the direct work proportion was not always accompanied by an increase in labour productivity. The scatter plots for the other activities, included in Appendix C.1, also showed similar result. The relationship between direct work proportion and CLP was then analyzed using the linear and nonlinear regression lines. A statistical significance test of the hypothesis that direct work is a significant predictor of the dependent variable (CLP) was conducted using the global F-test and a significance level of 0.05. The results, shown in Table 4.4, indicate that for all activities, the null hypothesis that direct work predicts the dependent variable (CLP) cannot be rejected. This finding is consistent with Handa and Abdalla's data analysis (1989), which determined an F-statistic value of 0.820 at p = 0.604.

The coefficient of determination, based on R^2 values, of the fitted regression lines shown in Table 4.4 indicated that the linear activity models for column concreting, slab concreting, wall concreting, box installation, piping, wire pulling, overlays, and shield installation were able to explain 20.6%, 2.9%, 3.5%, 5.6%, 0.8%, 0.1%, 10.0%, and 3.6%, respectively, of the variability in CLP, and all had low R^2 values. Similarly, the nonlinear activity models for column concreting, slab concreting, wall concreting, box installation, piping, wire pulling, overlays, and shield installation, were able to explain 24.3%, 3.4%, 16.3%, 8.6%, 3.7%, 2.7%, 13.3%, and 6.7%, respectively, of the variability in CLP (see Table 4.3); though improved, these R^2 values are again quite low. These results suggest that the relationship between productivity and direct work is complex, and an increase in direct work alone might not result in better labour productivity; however, the addition of other work sampling categories might provide a better result. This finding completely aligns with Thomas (1991), who concluded that direct work is not significantly correlated to productivity.

Table 4.4: CLP as Function of Direct Work Proportion: Model Summary

Activity	**Model**	**R^2**	**F**	**Sig.**
Column concreting	Linear	0.206	3.896	0.067
	Quadratic	0.243	2.246	0.143
Slab concreting	Linear	0.029	0.638	0.434
	Quadratic	0.034	0.348	0.710
Wall concreting	Linear	0.035	0.788	0.384
	Quadratic	0.163	2.039	0.155
Box installation	Linear	0.056	1.956	0.171
	Compound	0.086	3.119	0.087
	Growth	0.086	3.119	0.087
	Exponential	0.086	3.119	0.087
	Logistic	0.086	3.119	0.087
Piping	Linear	0.008	0.391	0.535
	Quadratic	0.037	0.894	0.416
Wire pulling	Linear	0.001	0.022	0.884
	Quadratic	0.027	0.410	0.667
Overlays	Linear	0.100	3.013	0.094
	Power	0.133	4.148	0.052
Shields installation	Linear	0.036	1.751	0.192
	Quadratic	0.067	1.659	0.202

4.4.2: Work Sampling Proportions versus CLP

The other assumption of activity models—that work sampling proportions (WSP) can predict CLP—was tested using the data collected for the eight activities shown in Table 4.1. Using general multiple regression analysis, activity models shown in Eq. (4.1) were developed for the eight activities, where Z = CLP, y_1 = Direct work, y_2 = Preparatory work, y_3 = Material handling, y_4 = Tools and equipment, y_5 = Waiting, y_6 = Travel, and y_7 = Personal:

$$Z = b_o + b_1 y_1 + b_2 y_2 + b_3 y_3 + b_4 y_4 + b_5 y_5 + b_6 y_6 + b_7 y_7 \quad (4.1)$$

The results of the correlation analysis, shown in Table 4.5 for column concreting activity, between the seven work sampling proportions (WSP) and CLP showed that some of the independent variables (WSP) are correlated to each other. The correlation analysis results for the other activities, included in Appendix C.2, also showed similar result.

Table 4.5: Pearson Correlation: Work Sampling Proportion with CLP for Column Concreting Activity

WSP	**Correlation coefficient of independent variables**							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.206 (0.427)	1						
y_3	0.392 (0.120)	-0.270 (0.295)	1					
y_4	-0.700 [a] (0.002)	0.241 (0.352)	−0.490 [a] (0.046)	1				
y_5	-0.618 [a] (0.008)	-0.492 [a] (0.045)	-0.409 (0.103)	0.238 (0.358)	1			
y_6	-0.552 [a] (0.022)	-0.298 (0.245)	-0.482 (0.050)	0.281 (0.358)	0.714 [a] (0.001)	1		
y_7	c (c)	c (c)	c (c)	c (c)	c (c)	c (c)	1	
Z	0.454 (0.067)	-0.398 (0.114)	-0.118 (0.651)	-0.109 (0.677)	-0.035 (0.895)	0.043 (0.871)	c (c)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test and *c* indicates that the value cannot be computed as the data instance values are constant.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Therefore, the magnitude of multicollinearity was checked using variance inflation factor, and for all activities the variance inflation factor was found to be greater than 10, indicating high multicollinearity (Keith 2015). Therefore, a stepwise regression analysis was performed for testing activity CLP models. Stepwise regression carries out multiple regressions a number of times, each time removing the weakest correlated independent variable, evaluated based on partial F-test values (Moselhi and Khan 2012).

Accordingly, the stepwise regression analysis on column concreting, slab concreting, wire pulling, overlays, and shield installation activities data showed that none of the independent variables (WS proportions) have a statistically significant relationship with the dependent variable (CLP). Similar analyses on wall concreting, box installation, and piping activities showed that only material handling, preparatory, and preparatory WS categories, respectively, had a statistically significant relationship with CLP.

Further diagnosis was carried out to identify highly influential data instances that could cause a large difference in the regression analysis results. The highly influential data instances were identified using influence measures, based on Cook's distance with a cut-off value of close to 1 (> 0.99) and centered leverage values with a cut-off value of $4\ /\ n$, where n is the total number of data instances (Keith 2015). Accordingly for wall concreting one data instance, for box installations activity one data instance, and for piping activity two data instances were found to be highly influential and could cause a large difference in the regression results. The stepwise regression analysis was repeated after removing the highly influential data instances for both wall concreting and box installation activities; however, the resulting models performance in terms of R^2 and $R^2_{\,adjusted}$ values did not show any improvement. Similar regression analysis was repeated for the piping activity, and three options were tested by removing: (1) the first influential data instance, (2) the second influential data instance, and (3) both influential data instances; the regression model based on the second option was selected as it showed an improvement in model performance. Finally, a check on the assumptions in multiple regressions was carried out for the three significant and diagnosed activity models. Accordingly, normality of the residuals was verified using standardized histograms and normality plots, and a lack of a pattern in the scatter plots of residuals was also confirmed. The results, shown in Table 4.3, summarize the model testing results for the selected activity models, which have only one independent variable as shown in Eqs. (4.2) to (4.4), where Z = CLP, y_2 = Preparatory work, and y_3 = Material handling.

$$Z_{Wall\ concreting} = 4.876 - 16.668\, y_3 \tag{4.2}$$

$$Z_{Box\ installation} = 2.539 - 5.467\, y_2 \tag{4.3}$$

$$Z_{Piping} = 5.250 - 11.025\, y_2 \tag{4.4}$$

However, the coefficient of determination, based on R^2 values, of the activity models for wall concreting, box installation, and piping activities showed that the models explained only 21.6%, 12.2%, and 12.9%, respectively, of the variability in CLP. As the coefficient of determination is low, the prediction error is expected to be high, and the usefulness of the activity models for predicting actual CLP is limited. These results suggest that the relationship between productivity and direct work is complex, and an increase in direct work alone might not result in better labour productivity. This finding completely aligns with the studies by Thomas (1991) and Josephson and Björkman (2013), which concluded that direct work is not significantly related to productivity. This study proves the limitations of activity models based on actual field data consistently collected from various projects and using accurate measures of CLP, standard WS categories, a strict data collection protocol, and a wide variety of activities ranging from commercial concreting of structural elements to boiler maintenance work in shutdowns. However, failures to find a significant relationship indicate that other factors or intermediate variables might have caused the failure to detect any significant relationship (Horman and Kenley 2005). The findings, thus, indicate that use of a system-based CLP model is necessary, as optimal solutions to any input-to-output conversion system can be found only if all the relevant parameters and components of the system are properly analyzed (Chang and Ibbs 2006). The research described in the following section presents a system CLP model that aims to advance the research framework of modeling CLP.

4.5: SYSTEM-BASED CLP MODELING APPROACH (*I–P–O* Relationship)

According to Chang and Ibbs (2006), organization operations and project performance have often been studied using system models. The present study develops the system model shown in Fig. 4.3. The system model consists of input (*I*), or independent, variables representing key parameters (factors and practices) influencing CLP; process (*P*) variables representing the seven work sampling proportions; and an output (*O*), or dependent, variable representing CLP. The process variables are treated as either mediator or moderator variables. In system models, *I* variables are assumed to cause an *O* variable, and this effect is represented as a direct effect (refer to Fig 4.3). Additionally, *P* variables can have either a mediating or moderating effect. In a mediation case, the *P* variables will have an indirect effect on the *O* variable, and the total effect of the variables *I* and *P* on *O* will be the sum of the direct and indirect effects. "Complete mediation" occurs when, as a result of the mediating effects of the *P* variables, the *I* variables no longer affect the *O* variable, and the direct effect of *I* on *O* is zero (Hayes 2013). In contrast, in a moderation case, the *P* variables alter the strength of the causal relationship between the *I* and dependent (*O*) variables and may amplify or even reverse the direct effect of *I* on *O* (Hayes 2013).

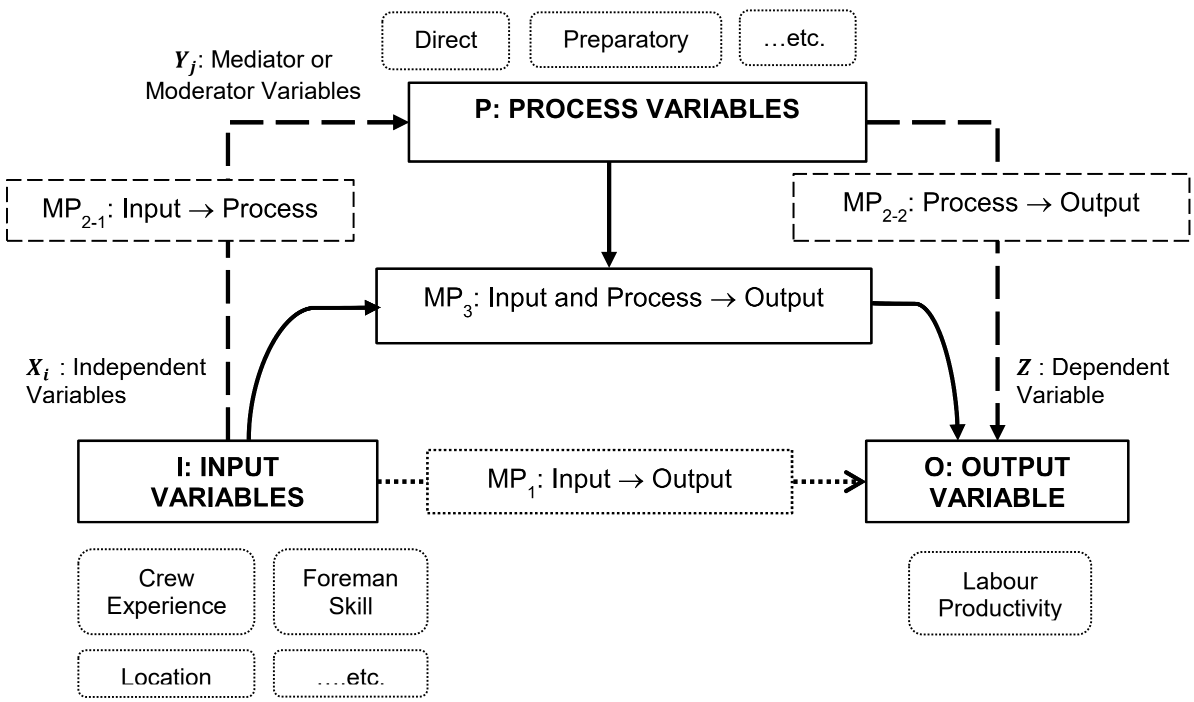

Figure 4.3: Modeling Paths for System Model of CLP

Thus, depending on the role of the *P* variables (work sampling proportions) in explaining the variability of CLP, the system model will constitute three paths (refer to Fig 4.3). The first path, represented as MP_1, is based on the *I*–*O* relationship and comprises the factor CLP model. The second path, represented as MP_2, is based on the *I*–*P*–*O* relationship and assumes that process variables have a mediating effect; it comprises a "mediated system" CLP model. In the mediated system CLP model, the assumption is based on complete mediation, where the *I* variables influence the *P* variables as mediator variables, which in turn influence the output or dependent variable (*O*). The third path, represented as MP_3, is based on the (*I* and *P*)–*O* relationship and assumes that *P* variables have a moderating effect; it comprises a "moderated system" CLP model. In the moderated system CLP model, the assumption is that the *P* variables, as moderator variables, affect the direction and strength of the relationship between the *I* and *O* variables.

4.5.1: Implementation Procedure

The following procedure was developed to test the proposed system CLP model. These steps indicate how to prepare the *I–P–O* data, develop the three path models, and test and select the most accurate system model of CLP:

1. Identify parameters influencing CLP and develop a measurement scheme. Also define labour utilization measures using work sampling proportions and establish an appropriate output measure for calculating CLP.
2. Select a labour-intensive activity and collect data on parameters, work sampling proportions, and CLP, which will formulate the input, process, and output variables, respectively, of the system model.
3. Establish model input features (F_1) based on the key parameters influencing CLP for the factor model, or model path MP_1, based on the *I–O* relationship.
4. Establish model input features (F_{2-1}) based on the key parameters influencing the process variables for the mediated system model, or model path MP_{2-1}, based on the *I–P* relationship. Also, establish model input features (F_{2-2}) based on the key process variables influencing the CLP variable for model path MP_{2-2}, based on the *P–O* relationship.
5. Establish model input features (F_3) based on the key parameters and process variables influencing CLP for the moderated system model, or model path MP_3, based on the (*I* and *P*)–*O* relationship.
6. Analyze the *I–O* relationship using the factor models in the MP_1 path.
7. Analyze the *I–P–O* relationship using the mediated system models in the MP_2 path, which includes MP_{2-1} and MP_{2-2} sub-paths.
8. Analyze the (*I* and *P*)–*O* relationship using the moderated system models in the MP_3 path.
9. Compare and contrast the overall accuracy of results of the three model paths and identify the most accurate system model.

4.6: System Model of CLP for Concreting Activity

The preceding system model and procedure were tested using the field data collected for this research. Analysis focused on concreting data collected from six projects (see Table 2.9). Data instances from the six structural elements were combined, which, compared to the other activities studied (also shown in Table 4.1), produced the largest data set with a total of 92 instances. Due to data size limitations, system models of the other activities were not developed. The results of the investigation of the three model paths for the *I–P–O* relationship for the concreting activity are presented here.

4.6.1: Establishment of System Model Features

Before testing the mediation and moderation effect assumptions presented above, feature extraction techniques like factor analysis are commonly applied in system modeling so as to reduce the input variable feature space (Hayes 2013). In this study, the 169 input parameters and 7 process variables result in a high-dimension *I–P–O* feature space, which must be reduced. The data-driven approach discussed in Section 3.4 for identifying key parameters was used to reduce the feature space. The most relevant features were selected using the Waikato Environment of Knowledge Analysis (WEKA) tool. The CfsSubsetEval algorithm was employed as the attribute subset evaluator with the BestFirst algorithm for the search method, as together, these provide the most efficient approach to correlation-based feature selection (Rodrigues et al. 2015). Key parameters, or features having high correlations with CLP but low inter-correlations, were identified for use in developing the system model.

The results of the feature selection process are shown in Table 4.6. The F_1 model input features for the factor model, or path MP_1, were selected from the *I–O* data; 15 of 106 features were selected. Similarly, the F_{2-1} model input features were selected for the first part of the mediated system model, or path MP_{2-1}, from the *I–P* data; 43 of 108 features were selected. Next, the F_{2-2} model input features were selected for the second part of the mediated system model, or path MP_{2-2}, from the *P–O* data; 4 of 7 features were selected. Finally, the F_3 model input features for the moderated system model, or path MP_3, were selected from the (*I* and *P*)–*O* data; 16 of 112 features were selected. Full descriptions of the selected features or variables, made up of key input and process parameters, are shown in Table 4.6.

Table 4.6: System Based CLP Models: Features Used for Model Development

CLP model	Model path	Feature selection F	S	S/F	Selected features or model variables, ID
Factor model	MP_1	106	15	13%	Craftsperson motivation$_{x11}$, craftsperson fatigue$_{x12}$, craftsperson trust in foreman$_{x13}$, level of absenteeism$_{x15}$, job security$_{x23}$, availability of tools$_{x37}$, change of foremen$_{x58}$, project work times$_{x96}$, owner staff on site$_{x97}$, approval of shop drawings and sample materials$_{x101}$, labour productivity measurement practice$_{x126}$, availability of communication devices$_{x143}$, concrete placement technique$_{x45}$, structural element$_{x46}$, **construction labour productivity$_z$**
Mediated system model	MP_{2-1}	112	43	39%	Project type$_{x70}$, craftsperson education$_{x2}$, craftsperson on-job training$_{x3}$, crew composition$_{x5}$, crew experience (seniority)$_{x6}$, co-operation among craftsperson$_{x8}$, treatment of craftsperson by foreman$_{x10}$, craftsperson motivation$_{x11}$, craftsperson fatigue$_{x12}$, level of absenteeism$_{x18}$, discontinuity in crew makeup$_{x17}$, fairness of work assignment$_{x19}$, crew flexibility$_{x21}$, availability of craftsperson$_{x24}$, quality of task materials$_{x26}$, material unloading practices$_{x31}$, material movement practices (horizontal)$_{x32}$, equipment breakdown$_{x36}$, availability of tools$_{x37}$, availability of extension cords$_{x40}$, complexity of task$_{x41}$, level of rework$_{x42}$, frequency of rework$_{x43}$, task change orders$_{x44}$, working condition (noise)$_{x47}$, congestion of work area$_{x50}$, foreman skill and responsibility$_{x56}$, fairness in performance review of crew by foreman$_{x57}$, change of foremen$_{x58}$, approval for building permit$_{x67}$, queue time to access site $_{x91}$, support and administrative staff$_{x104}$, availability of labour$_{x112}$, project cost control practice$_{x125}$, out of sequence inspection$_{x129}$, safety incident investigation$_{x137}$, communication between different trades$_{x142}$, project team development$_{x145}$, sorting of waste materials$_{x147}$, oil price fluctuation$_{x167}$, natural gas price$_{x168}$, concrete placement technique$_{x45}$, type of structural element$_{x46}$, **direct work$_{y1}$**
	MP_{2-2}	8	5	57%	*Direct work$_{y1}$, preparatory work$_{y2}$, travel$_{y6}$, personal$_{y7}$,* **construction labour productivity$_z$**
Moderated system model	MP_3	113	17	14%	Craftsperson motivation$_{x11}$, craftsperson fatigue$_{x12}$, craftsperson trust in foreman$_{x13}$, level of absenteeism$_{x15}$, job security$_{x23}$, availability of tools$_{x37}$, change of foremen$_{x58}$, project work times$_{x96}$, owner staff on site$_{x97}$, approval of shop drawings and sample materials$_{x101}$, labour productivity measurement practice$_{x126}$, availability of communication devices$_{x143}$, concrete placement technique$_{x45}$, structural element$_{x46}$, *direct work$_{y1}$*, *personal $_{y7}$*, **construction labour productivity$_z$**

Note: *F* = total feature space based on collected parameters and/or work sampling categories, *S* = selected feature space after running the feature selection algorithm. Features in italics indicate process variables, and features in bold indicate the target variable.

4.6.2: Development and Evaluation of System Models

Tests on the mediation and moderation effect assumptions are traditionally carried out using regression analysis approaches like structural equation modeling (Hayes 2013). According to Green (1991), the desired sample size (DS) for testing multiple regression is DS > 50 + 8F (where F is the number of independent variables) and the absolute minimum size is 30. This study examines a total of 169 independent input and 7 process variables; even if only 15% of these variables were selected using the feature selection process, the desired sample size required for regression analysis would be 261 (i.e., 50 + 8 × 176 × 0.15).

Collecting such a large number of data instances is not realistic in CLP studies, as this would require extensive data collection over an extended period of time—an expensive, time-consuming endeavor. Therefore, as regression analysis is not realistic, this study use artificial intelligence (AI) techniques like neural networks and fuzzy rule-based models. Such techniques are able to deal with high dimension feature space with limited data, so by adopting them, the limitations associated with regression analysis are overcome.

In this way, the mediation and moderation effect assumptions were tested by developing AI technique-based models and evaluating which model and path provided the most accurate results. As testing the system model assumption using only one modeling approach could be misleading, three AI-based approaches were studied: (1) neural networks (NN); (2) a fuzzy inference system (FIS_m) based on Mamdani fuzzy rule-base; and (3) a fuzzy inference system (FIS_s) based on Sugeno fuzzy rule-base. AI models were developed for each of these approaches in the MATLAB® 2014 environment, and used identical model architectures for the three paths. The NN models used a feedforward backpropagation learning algorithm with a hidden layer, 10 neurons, a hyperbolic sigmoid transfer function, and a learning rate of 0.01—all of which are the default model parameters in MATLAB's NN Tool Box.

Fuzzy inference or rule-based models use fuzzy sets and if-then rules with condition and conclusion parts. In the case of Mamdani fuzzy rule-based models, the conclusion is represented as a fuzzy set and defuzzification is employed to obtain a crisp output value. In Sugeno fuzzy rule-base models, the conclusion is represented using a function (Pedrycz and Gomide 2007). Both Mamdani fuzzy rule-based (FIS_m) and Sugeno fuzzy rule-based (FIS_s) models were developed using the MATLAB "genfis3 (Mamdani)" and "genfis3 (Sugeno)" functions, respectively. The functions generate the respective fuzzy inference models using Fuzzy C-Means (FCM) clustering technique and approximated membership functions of each cluster as a Gaussian membership functions for all fuzzy sets. The FCM algorithm extracts the rules that model the behavior of the data and develops the fuzzy sets membership functions in the rule condition and conclusion for Mamdani fuzzy rule-based models, and only in the rule condition for Sugeno fuzzy rule-based models. The number of rules was determined using the subtractive clustering method with radii of 0.5 and the minimum and maximum values of the variables included in the model input feature space and output variable as limiting boundaries.

The Mamdani fuzzy rule-based model consists of four parts: fuzzification, implication, aggregation, and defuzzification. In the fuzzification stage, the model calculates the degree of membership for each fuzzy set based on the value of each variable. Then, for each rule, a single value is calculated by applying a fuzzy operator to the membership values of each variable. The MIN (minimum) fuzzy operator was used to combine the different parts of the conditions of the rules and also for implication. The MAX (maximum) operator was used for rule aggregation, and the centroid method for defuzzification. These fuzzy operators are the default model parameters in MATALB's Fuzzy Tool Box for Mamdani fuzzy rule-based models. The Sugeno fuzzy rule-based model consists of three parts: fuzzification, implication, and aggregation. Fuzzification is achieved similarly to the Mamdani model. The PROD (product) fuzzy operator was used to combine the different parts of the conditions of the rules and for the implication method. The MAX operator was used for rule aggregation. These fuzzy operators are the recommended operators in Sugeno fuzzy rule-based models (Pedrycz and Gomide 2007).

To test and select the most accurate system model, this study employed a number of model performance metrics. For each of the three model paths, AI-based models were developed using NN, Mamdani fuzzy rule-based (FIS_m), and Sugeno fuzzy rule-based (FIS_s) architectures. For each AI model, the mean prediction error (E_i) and standard deviation of prediction errors (STD_i) were used to evaluate accuracy. As the objective is to evaluate which model path is most accurate, for each path, the accuracy results of the three different AI models were combined to get an overall accuracy evaluation of the model path. The combination of the results of the AI models will ensure the path selection is not dependent on one specific model architecture but rather on the strength of the path in providing a better explanation of CLP and its variance. Timeliness, precision, repeatability, and accuracy performance metrics were used to determine the overall accuracy of a given model path (Zemouri et al. 2010). Suppose *n* represents the total number of data instances for the activity under investigation, *i* represents each of the AI models in the three paths (for this study *i*=9), and MP_k represents a given model path where *MP* represents a model path and *k* represents the number of model paths (for this study *k*=*3*). For every model, the mean prediction error and standard deviation of errors are computed using Eq. (4.5) and Eq. (4.6), where t_j is the target value for the j^{th} data instance and y_j is the corresponding predicted value.

$$E_i = \frac{1}{n}\sum_{j=1}^{n}\sqrt{(t_j - y_j)^2} \tag{4.5}$$

$$STD_i = \sqrt{\sum_{j=1}^{n}\frac{1}{n}(E_i - y_j)^2} \tag{4.6}$$

Timeliness of a given model path (MP_k) is measured by calculating the mean of the mean of prediction errors of the three AI models developed for the model path (Eq. 4.7). Precision of a given model path is measured by calculating the mean of the standard deviations of prediction errors of the three AI models developed for the model path (Eq. 4.8). Repeatability of a given model path is measured by averaging the standard deviation of both the mean of prediction errors and standard deviations of prediction errors of the three AI models developed for the model path (Eq. 4.9). In Eq.s (4.5) through (4.7), E_{ik} is the mean prediction error of each i^{th} AI model in the k^{th} model path, and STD_{ik} is the standard deviation of error of each i^{th} AI model in the k^{th} model path.

$$Timeliness = \overline{E_k} = \frac{1}{k}\sum_{i=1}^{k} E_{ik} \tag{4.7}$$

$$Precision = \overline{STD_k} = \frac{1}{k}\sum_{i=1}^{k} STD_{ik} \tag{4.8}$$

$$Repeatability = \frac{\sigma\ (E_{ik}) + \sigma\ (STD_{ik})}{2} \tag{4.9}$$

Accuracy measures the global accuracy of the prediction path and is calculated as in Eq. (4.10). The accuracy measure of the three model paths is compared, and the model path that gives the highest accuracy is considered to be the appropriate system model.

$$Accuracy = \frac{1}{Timeliness + Precision + Repeatability} \tag{4.10}$$

The perfect score for timeliness, precision, and repeatability performance measures is achieved when the values are equal to zero, and a larger value of accuracy performance measure indicates a greater confidence in the predictive capability of the model path in predicting CLP (Zemouri et al. 2010).

4.6.3: Results and Discussion

The factor model based on the *I*–*O* relationship, or model path MP_1, was analyzed, and three AI models were developed using the selected 15 features shown in Table 4.6. The performance measures of the factor model shown in Table 4.7 indicated that the overall accuracy of the model path was 0.2768. The *I*–*P*–*O* relationship, or model path MP_2, was analyzed using the mediated system model; three AI models were developed using the selected 43 features for model path MP_{2-1} and 4 features for model path MP_{2-2}, both shown in Table 4.6. The performance measures of the mediated system model, shown in Table 4.7, indicated that the overall accuracy of the model path was 0.2609. The (*I* and *P*)–*O* relationship, or model path MP_3, was analyzed using the moderated system model; three AI models were developed using the selected 17 features shown in Table 4.6. The performance measures of the moderated system model shown in Table 4.7 indicated that the overall accuracy of the model path was 0.3042.

Table 4.7: System Based CLP Models Performance Analysis

CLP model	Path	Model type	E_i	STD_i	Timeliness	Precision	Repeatability	Accuracy
Factor model	MP_1	NN	1.418	1.831	1.477	1.983	0.152	0.2768
		FIS_m	1.645	2.243				
		FIS_s	1.369	1.875				
Mediated system model	MP_{2-1} MP_{2-2}	NN	1.861	2.019	1.765	2.015	0.053	0.2609
		FIS_m	1.809	2.020				
		FIS_s	1.626	2.006				
Moderated system model	MP_3	NN	1.045	1.741	1.260	1.881	0.147	0.3042
		FIS_m	1.340	2.071				
		FIS_s	1.396	1.829				

With an accuracy value of 0.3042, of the factor (MP_1), mediated system (MP_2), and moderated system (MP_3) models shown in Table 6, the moderated system model was found to be the most accurate, proving that process variables (*P*) have a moderation effect is accurate. The mean of prediction errors of the three AI models for the moderated model path were also consistently lower than the other two model paths. This finding implies that process variables or work sampling proportions, do not directly affect CLP but rather strengthen the influence of the parameters on CLP, as the moderated system model had 9.9% accuracy improvement over the factor model due to the moderation effect of the process variables. The moderated system model also had 16.7% accuracy improvement over the mediated system model. Timeliness performance measures also indicated that the moderated system model had the lowest timeliness value (1.260), indicating that this model has a higher chance of accurately predicting CLP compared to the factor model and mediated system model. Precision performance measures also indicated that, with a precision value of 1.882, the moderated system model performs better than the other two, as the predicted values from the moderated system model are less dispersed and more specific. However, with a repeatability value of 0.053, the mediated system model showed better repeatability performance compared to the other two models.

The comparison of the factor and mediated models indicated that the factor model had better accuracy than the mediated system model, which implies that the assumption that input variables influence the process variable, which in turn influences CLP, is not accurate and a causal relationship between input variables (key parameters) and CLP does exist. Thus, CLP optimization can be achieved by adjusting the key factors and practices influencing it. Additionally, the moderated system model shows that work sampling proportions do not directly affect CLP, but strengthen the influence of the parameters on CLP. Therefore, the results imply that focusing improvement efforts on key factors and practices identified for model path MP_3, as shown in Table 4.6, would lead to improved CLP—but only in combination with appropriate utilization of labour time characterized by higher direct work and lower personal work time proportions. Based on these findings, the following factors and practices should be considered in order to improve CLP: availability of tools, use of concrete pumps for placement, labour productivity measurement practice, craftsperson trust in foreman, level of absenteeism, craftsperson fatigue, and availability of communication devices. Additionally, the findings suggest that direct work and personal work proportions should be considered during execution of concreting (concrete placement) operations. Lastly, the accuracy of the moderated system model presents a novel approach to how work sampling proportions can be integrated with parameters (factors and practices) in future CLP modeling and analysis studies. The success of this approach supports the recommendation that future CLP studies should carry out work sampling studies in conjunction with documentation of key parameters to accurately measure productivity.

4.7: Chapter Summary

The construction industry is constantly searching for ways to improve construction labour productivity (CLP), but until now the industry has lacked crew-level CLP models capable of explaining which parameters (factors and practices) cause productivity to change. As a result, the industry has been unable to propose CLP improvement strategies that accurately reflect the construction environment, where tradespeople work together in crew units. This chapter presents a research methodology that integrates a variety of traditional data collection methods used in CLP studies—namely: factors and practices documentation, work sampling (WS) studies, foreman delay surveys, craftsman questionnaires, and productivity measurement—to formulate input, process, and output variables for use in CLP system modeling.

Using extensive field data collected from 11 construction projects across Alberta, Canada, this chapter examined the validity of the existing activity and factor modeling approaches. It tested the validity of the activity models by investigating the relationship between CLP and seven work sampling proportions for 8 activities: concreting for columns, slabs, and walls; electrical box installation, piping, and wire pulling; and shutdown overlays and shield installations. The investigation showed that direct work proportions are not significantly correlated to CLP and accurate prediction of CLP is not possible with either linear or nonlinear regression models. No significant correlations between the proportion of direct work and CLP could be observed, so direct work proportions cannot be used as surrogate measures of CLP; furthermore, activity models are not able to explain the variability of CLP and need additional explanatory parameters to improve their predictive capability.

This chapter proposes a novel system model approach for improved prediction of CLP using input variables made up of key influencing parameters in conjunction with process variables made up of work sampling proportions. Three models, namely, a factor model, mediated system model, and moderated system model, were formulated and evaluated. The approach used the correlation-based feature selector (CFS) algorithm to reduce the high dimension of the independent variables, and artificial intelligence (AI) techniques based on neural networks, Mamdani fuzzy rule-base (FIS_m) models, and Sugeno fuzzy rule-base (FIS_s) models to formulate, test, and determine the most accurate system model. The analysis results showed that the moderated system model is the most accurate; it also had the best performance for timeliness and precision measures. The moderated system model proves that process variables have a moderating effect on factor models. The analysis results further indicate: (1) a causal relationship between parameters and CLP exists, and optimization of CLP can be achieved by adjusting the parameters or key factors and practices influencing it; (2) work sampling proportions do not directly affect CLP but rather strengthen the influence of the parameters on CLP; and (3) work sampling proportions can be integrated with parameters in future CLP modeling and analysis studies following the presented methodology. In the next chapter, using the findings of the moderated system model approach and the context-specific nature of the key parameters influencing CLP, a framework for the development of context-specific CLP models based on fuzzy inference systems (FIS) is presented.

4.8: REFERENCES

Awad, A., and Fayek, A. Robinson. (2012). “A decision support system for contractor prequalification for surety bonding.” *Autom. Constr.*, 21, 89–98.

Chang, A., and Ibbs, W. (2006). “System model for analyzing design productivity.” *J. Manage. Eng.*, 22(1), 27–34.

CII (Construction Industry Institute). (2010). *Guide to activity analysis*, Univ. of Texas at Austin, Austin, Texas, US.

Fayek, A. Robinson, and Tsehayae, A. A. (2012). "Modeling construction labour productivity using fuzzy logic and exploring the use of fuzzy hybrid techniques." *Fuzzy Information Processing Society (NAFIPS), 2012 Annual Meeting of the North American,* IEEE, Berkeley, California, US, 1–6.

Fulford, R., and Standing, C. (2014). “Construction industry productivity and the potential for collaborative practice.” *Int. J. Project Manage.*, 32(2), 315–326.

Gerek, İ. H., Erdis, E., Mistikoglu, G., and Usmen, M. (2015). “Modelling masonry crew productivity using two artificial neural network techniques.” *J. Civ. Eng. Manage.*, 21(2), 231–238.

Gouett, M., Haas, C., Goodrum, P., and Caldas, C. (2011). “Activity analysis for direct-work rate improvement in construction.” *J. Cons. Eng. and Mgmt.*, 137(12), 1117–1124.

Green, S. (1991). “How many subjects does it take to do a regression analysis?” *Multivar. Behav. Res.*, 26, 499–510.

Guyon, I., and Elisseeff, A. (2003). “An introduction of variable and feature selection.” *J. Mach. Learn. Res.*, 3, 1157–1182.

Handa, V., and Abdalla, O. (1989). “Forecasting productivity by work sampling.” *Constr. Manage. Econ.*, 7(1), 19–28.

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press, New York, US.

Horman, M. J., and Kenley, R. (2005). "Quantifying levels of wasted time in construction with meta-analysis." *J. Constr. Eng. Manage*.,131 (1), 52-61.

Josephson, P., and Björkman, L. (2013). “Why do work sampling studies in construction? The case of plumbing work in Scandinavia.” *Eng. Construct. Architect. Manage.*, 20(6), 589–603.

Keith, T. Z. (2015*). Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*, 2nd Ed., Routledge, New York, US.

Moselhi, O., and Khan, Z. (2012). “Significance ranking of parameters impacting construction labor productivity.” *J. Cons. Innov.: Info., Proc., Manage*., 12(3), 272–296.

Nasirzadeh, F., and Nojedehi, P. (2013). "Dynamic modeling of labor productivity in construction projects." *Int. J. Project Manage.*, 31(6), 903–911.

Oduba, A. O. (2002). “Predicting industrial construction productivity using fuzzy expert systems.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Pedrycz, W., and Gomide, F. (2007). *Fuzzy systems engineering: Toward human-centric computing.* John Wiley and Sons, Hoboken, New Jersey, US.

Rodrigues, D., Prata, D., and Silva, M. (2015). “Exploring social data to understand child labor.” *Int. J. Social Sci. Humanity*, 5, 29–33.

Shahtaheri, M., Nasir, H., and Haas, C. T. (2014). "Setting baseline rates for on-site work categories in the construction industry." *J. Constr. Eng. Manage.,*141(5), 04014097.

Sheppard, B., and Lewicki, R. J. (1987). “Toward general principles of managerial fairness.” *Soc. Just. Res.*, 1(2), 161–176.

Silva, S. P., and Ruwanpura, J. Y. (2006). “Concreting productivity improvements in Alberta on commercial construction projects.” *Proc., 1st Int. Constr. Specialty Conf.*, CSCE, Calgary, Alberta, Canada.

Thomas, H. (1991). “Labor productivity and work sampling: The bottom line.” *J. Constr. Eng. Manage*., 117(3), 423–444.

Thomas, H., Maloney, W. F., Horner, R. M., Smith, G., V.K., H. and Sanders, S. (1990). “Modeling construction labor productivity.” *J. Constr. Eng. Manage.*, 116(4), 705–726.

Thomas, H., and Sakarcan, A. (1994). “Forecasting labor productivity using factor Model.” *J. Constr. Eng. Manage.*, 120(1), 228–239.

Tsehayae, A. A., and Fayek, A.R. (2014). “Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects.” *Can. J. Civ. Eng*., 41, 878–891.

Yi, W., and Chan, A. P. (2014). “Critical review of labor productivity research in construction journals.” *J. Manage. Eng.*, 30(2), 214–225.

Zemouri, R., Gouriveau, R. and Zerhouni, N. (2010). “Defining and applying prediction performance metrics on a recurrent NARX time series model.” *Neurocomputing*, 73(13–15), 2506–2521.

Chapter 5: Developing and Optimizing Context-Specific CLP Models[5]

5.1: INTRODUCTION AND BACKGROUND

Construction labour productivity (CLP) is of critical importance to the construction industry, as it directly affects the profitability and competitiveness of construction companies (Song and AbouRizk 2008), and it is therefore a frequently researched topic. Nevertheless, labour productivity continues to be a major source of construction risk and exhibits the highest variability among construction resources (Tsehayae and Fayek 2014). CLP is an efficiency measure of an activity-level open system that deals with the process of converting inputs (material, information, etc.) to outputs (project components) using labour as the chief transformation mechanism. In this study, CLP is defined as the ratio of units of output—in terms of installed quantity—to units of input—in terms of total labour work-hours—with the attainment of higher CLP values being the objective of CLP systems. It is important to note that the environment of CLP systems is more complex and unpredictable than the construction process itself, causing a number of parameters to either directly or indirectly influence CLP.

Several CLP studies have identified numerous parameters that influence CLP (Thomas et al. 1990; Fayek and Oduba 2005; Song and AbouRizk 2008; Oral et al. 2012; Tsehayae and Fayek 2014; Gerek et al. 2014), and CLP models for analyzing the impact of the influencing parameters on CLP have been developed using a variety of modeling techniques (Yi and Chan 2014). Overall, the tested CLP models can be categorized as either factor or activity models. Factor models relate different input variables to labour productivity. These input variables are made up of key influencing parameters (factors and practices) such as crew size, weather conditions, etc. Activity models relate process variables, in terms of work sampling proportions, to labour productivity. Notably, the identified influencing parameters and the associated CLP models were all context-dependent, as the identified parameters and their degree of impact on CLP varied from project to project (Tsehayae and Fayek 2014; Gerek et al. 2014). However, only a few CLP studies explicitly defined the context of the CLP modeling processes. In a study

[5] Parts of this chapter have been submitted for publication in Journal of Construction Engineering and Management, JCEM, 33 manuscript pages, submitted June 30, 2015; and have been published in the Proceedings, NAFIPS Annual Meeting 2012, Berkeley, California, US, August 6-8, pp. 1-6.

to model the effect of delay time on concrete placement productivity, context was used during the data collection stage as a control variable specifying building type, type of equipment and method, project location, type of contract, and union type attributes so as ensure proper comparison among studied projects (Christian and Hachey 1995). Thomas and Raynar (1997) quantified the effect of scheduled overtime on CLP by combining data from four projects; they collected project features based on type of project, approximate cost, and approximate planned duration attributes in order to formulate the project conditions under which the construction work was done. Fayek and Oduba (2005), employed fuzzy inference systems to predict the CLP of pipe rigging and welding activities, and clearly defined context as a set of multi-leveled factors whose values are fixed for a given project scenario and/or activity. The authors used context to categorize activities and formulate the membership functions (MFs) of the fuzzy sets in the CLP model. The context attributes were established at two levels: project-level context variables included project location, year of construction, client, contract type, project type, and season of construction, while activity-level context variables included material type and weld type (applicable to welding activities only). Thus, most of the developed CLP models were focused either on the impact of a few selected influencing factors or on the advancement of the state of art of modeling techniques. Most models overlooked the role of context in CLP model development. Additionally, the development of accurate and interpretable CLP models has been a challenge due the complex variability of CLP and the limited data availability to study CLP under various contexts; thus, studies have focused on the use of artificial intelligence techniques like neural networks and fuzzy inference systems to model CLP (Oral et al. 2012; Fayek and Oduba 2005). Fuzzy inference systems (FISs) are based on fuzzy set theory and if-then rules, and have proved effective tools for solving engineering problems in biomedicine, robotics, pattern recognition, image processing, and control application areas (Botta 2008). The use of FISs has also been gaining widespread attention in construction research (Chan et al. 2009). However, there have been few applications of FISs in CLP modeling, and the few studies using FISs were limited in that they did not develop MFs and if-then rules from data (Mao 1999; Fayek and Oduba 2005). Additionally, FIS inability to learn from data and develop and optimize system parameters is a major limitation. Thus, hybridizing FIS through combination with other artificial intelligence techniques such genetic algorithms (GA) have been tested, resulting in improved learning capabilities (Awad and Fayek 2013).

This chapter provides a methodology that addresses CLP modelling challenges using a hybrid approach that incorporates fuzzy inference systems developed using a data-driven fuzzy clustering technique combined with a GA-based optimization process. The development of the CLP models begins with identifying, classifying, quantifying, and documenting the parameters influencing CLP. Past studies used either input variables made up of key influencing parameters (factors and practices) for developing factor CLP models or process variables made up of key work sampling proportions for developing activity CLP models. However, previous studies have not succeeded in developing an integrated system approach that captures the overall relationship between both input and process variables and CLP. To address this gap, a novel system model approach for improved prediction of CLP using input variables made up of key influencing parameters in conjunction with process variables made up of work sampling proportions was developed as discussed in Section 4.6, The system analysis results showed that the moderated system model is the most accurate, demonstrating that process variables have a moderating effect on factor models and concluded that: (1) a causal relationship between input variables (key factors and practices) and CLP exists; (2) work sampling proportions (process variables) do not directly affect CLP, instead they strengthen the influence of the input variables on CLP; and (3) work sampling proportions can be integrated with influencing factors and practices resulting in a moderated input variables for future CLP modeling and analysis studies. Thus, by adding the process variables to the list of factors and practices, the resulting moderated system model will improve the accuracy of predicting CLP as compared to using only the factors and practices in predicting CLP. Following these findings, the CLP influencing parameters in this study were formulated by adding process variables (work sampling proportions) to the input variables (factors and practices). Therefore, the 169 previously developed CLP parameters, which are comprised of various objective and subjective factors and practices (e.g., crew size, complexity of task, congestion of work area, safety training, etc.), are combined with 7 process efficiency measures based on work sampling proportions (direct work, preparatory work, tools and equipment, material handling, waiting, travel, and personal) and were used to define the CLP system and its environment. Accordingly, the quantification process, discussed in Chapter 2, to formulate appropriate measurement scales and facilitate field data documentation of the 176 influencing parameters has resulted in a total of 321 measurable variables. The variables were then used to investigate the following objectives of this chapter: (1) to apply an explicit operational definition of context for CLP model development and formulate context attributes, (2) to apply the operational definition of context to classify the studied construction projects based on context similarity, and study the complex variability of CLP using context-specific models addressing each of the identified unique contexts, (3) to investigate the effect of context in CLP modeling by comparing the performance of context-specific models against the generic CLP model, (4) to illustrate the application of a data-driven fuzzy clustering technique in the development of an interpretable FIS, and (5) to address the FIS's major limitation, namely, the inability to learn from data, by using a GA-based optimization process, and validate the FIS models using a leave-one-out strategy.

This chapter begins by reviewing the concept of context and formulating context attributes based on the operational definition for application in the field of CLP modeling. Next, the chapter discusses the procedure for developing context-specific CLP models and, using field data collected for concreting activity, applies the procedure to develop, optimize, and validate a series of context-specific CLP models. Subsequently, it performs a comparative assessment of the context-specific models and discusses the role of context in CLP modeling. Finally, conclusions are presented.

5.2: Defining Context for CLP Modeling

Computing fields like cognitive psychology, natural language processing, business, philosophy, linguistics, artificial intelligence, database integration, communication, and education sciences have vastly explored and applied the concept of context (Bazire and Brézillon 2005). Context has been recently applied in the construction domain to determine the meaning of information items coming from different databases used by architecture, engineering, and construction firms participating in the same construction projects (Zhu 2005). For real-world engineering applications, context is quite useful, as it restricts the state of a problem's space within which an effective representation can be developed (Brézillon and Pomerol 2001); context is therefore considered static in the modeling of real-world problems (Turner 2014).

However, establishing a complete description of context remains a challenge (Bazire and Brézillon 2005). In CLP modeling, the aim is to define context at a level that will enable the effective modeling of the construction process. Accordingly, this chapter relies on the following operational definition of context: context is what constrains the main elements of a CLP model without intervening in the model development process explicitly (Bazire and Brézillon 2005). Therefore, the context of a CLP model can be represented using a set of context attributes that define the properties of the main elements of a CLP model: (i) user of the developed CLP model, (ii) model developer, (iii) CLP model, and (iv) environment of the studied CLP system. Context attributes have been generated using the 5W1H questions approach: Who? What? Where? When? Why? and How? (Jang and Woo 2003). From the list of 169 factors and practices shown in Table 2.1, the static factors and practices providing the best answers to the "who," "what," "where," "when," and "why" questions were identified and mapped to the 5W context attributes. Table 5.1 shows a list of context attributes for CLP modeling together with example values.

The "who" aspect of context characterizes the conditions of the resource (labour, material, equipment, etc.) directly involved in the construction process, and the "what" addresses the CLP conversion process under investigation. The "where" and "when" aspects represent the location and time-specific properties, respectively, of the CLP model. The "why" aspect of context represents the intentions of the CLP system study (e.g., prediction and optimization of CLP, identifying effect of important factors like change order, etc.) and the expected user of the developed CLP model (e.g., cost estimators, project planners, researchers). The “how” aspect of context addresses the modeling methodologies adopted for identifying influencing parameters, quantifying parameters, collecting data, applying the modeling technique, and validating the models.

Table 5.1: Context Attributes for CLP Modeling with Example Values for Project 1 of Context 1

Project		**Context attributes**
Mixed-use office and staff facility project	WHO	Labour and crew [multiskilling of crew (civil), open shop]; Materials and consumables [policy on material management (yes), material order tracking system (no)]; Project owner team [owner staff on site (none), supervision from owner or representative (low), owner's primary driver (safety and cost)]; Project contractor team [foreman experience (6 months), superintendent experience (32 years) and project management experience (6 years)]; Contractor organization [organization's principal project type (commercial, institutional), organization year in industry (over 100 years), annual turnover , annual employee turnover, organizational structure (matrix), project load]; National [*foreign workers use* (3)]
	WHAT	Activity property [activity type (concrete pouring), total work volume (2500m^3)]; Engineering and instructions [use of standard specification (yes), *quality of specifications* (4), *quality of drawings* (4)]; Project delivery and contract type [project delivery system (design build), contract type (reimbursable), level of fast tracking (0%)]; Project nature [project type (commercial), project size, *project complexity* (3), project organization structure (line), project percent complete (5%)]; Project condition [*site layout* (4), *unloading or laydown area* (4)]; Project management practices [initial planning (*front end planning* (3), *constructability review* (2)), scope (*definition* (4), *verification* (2)), time (criticality of project schedule (80%)), cost (tracking system for labour productivity (none)), procurement (trade subcontracting (65%)), safety (use of safety officer (yes), daily job hazard assessment (yes), *safety practice implementation* (5)), and environment (environmental rating of project (LEED rated)]
	WHERE	Activity location [cover from weather effect (no)]; Project [location (Edmonton region)]; Provincial or state properties [name (Alberta), economy ($312 billion), total number of similar projects within province (105), unemployment rate (2.5%), labour strikes (none), income tax (25%), total expenditure in construction projects ($23 billion)]; National [name (Canada), *political system stability* (4), population (34.48 Million), interest rate (3%), inflation rate (2.08%)]
	WHEN	Year of CLP study data collection (2012), Global attributes [global economic outlook (1.5% real GDP growth)]
	WHY	Objectives of the CLP modeling study [prediction of CLP (yes), identifying effect of single factors like change order, rework, etc. (no)]; Expected users of developed CLP model [cost estimators (yes), project planners (yes), academic researchers (yes)]
	HOW	Modeling methodology [parameter identification (survey and data), data collection (field observation), modeling technique (fuzzy inference system), and validation method (leave-one-out strategy)]; Model developer [academic researcher (yes), productivity improvement professional (no)]

Note: Italicised context attributes are measured based on subjective 1-5 predetermined ratings.

The previous CLP modeling endeavors discussed above lack a clear and explicit representation of context and fail to consider context in the development of the CLP models. Furthermore, the knowledge captured by the models is not comprehensive, as model developers have examined the impact of only selected factors on very few activities (Yi and Chan 2014). Context plays an essential role in CLP research, as it defines in which scenarios the findings of the CLP models are applicable. For context-sensitive problems like CLP studies, it has been recommended that models are developed incrementally, when needed, and for a specific context of use (Green et al. 2009; Pedrycz et al. 2012; Botta 2008). Therefore, the research described in the following sections presents the development of a series of context-specific CLP models to improve understanding of the role context plays in CLP studies.

5.3: Fuzzy Inference Systems

Due to the numerous and varied input variables influencing construction labour productivity (CLP) and a lack of understanding of their effects, previous CLP studies have often focused on the relationship between limited and mostly objective parameters and the achieved productivity. However, the same reasons that make CLP notoriously difficult to predict—complexity, limited data availability, and a large number of subjective parameters that result in an uncertainty insufficiently resolved by statistical modeling and probability theory—make it an exceptional target for hybrid fuzzy modeling. The context-specific models are thus developed using FISs, the most widely used and central architecture in fuzzy modeling (Pedrycz and Gomide 2007). This approach has the capability to deal with CLP's large number of subjective variables by means of fuzzy partitions represented by linguistic terms, and to model the complexity of CLP using if-then rule base, which can be developed using limited data. The fuzzy sets representing the linguistic terms in the condition and conclusion parts of the if-then rules are characterized by their MFs, which numerically represent the degree to which an element belongs to a fuzzy set and fits the concept expressed by the linguistic term. Gaussian MFs have been recommended for their smoothness, having nonzero values at all points, possessing interpretability, and being suitable for optimization (Pedrycz and Gomide 2007). Nevertheless, few CLP studies have used fuzzy inference systems. A CLP model for concrete wall formwork was developed using FIS, but the MFs were developed without using data or experts and the if-then rules were developed using inconsistent historical data that did not include the necessary information for some variables like crew size (Mao 1999). In another case—fuzzy inference CLP models for industrial pipe rigging and welding activities—the MFs were developed using only two industry experts, and the rules were developed using simple logical reasoning rather than data (Fayek and Oduba 2005).

The quality of a linguistic fuzzy inference system is guided with two contradictory requirements: accuracy and interpretability (Cordón 2011). In contrast to the criterion of accuracy, which is quantified using measures like the root-mean-square error (RMSE), the interpretability of fuzzy rules is more difficult to describe and deals with the complexity and semantics of the rule-base and the fuzzy partition (Gacto et al. 2011). The interpretability of FIS depends on several aspects: the model structure, the number of rules, the number of input variables, the number of linguistic terms used to characterize the input and output variables, and the shape and overlap of the fuzzy sets used to represent the linguistic terms (Gacto et al. 2011). Accordingly, it is advisable to keep the number of linguistic terms and rules as small as possible, include only few key input variables, and use linguistic terms that are intuitively comprehensible, so as reduce the complexity of the rule-base and fuzzy partition and improve interpretability of the FIS (Gacto et al. 2011). Thus, in FIS modeling, feature selection is first carried out to reduce the number of input variables by determining the most important input variables (Ahmad et al. 2012).

The MFs and fuzzy if-then rules of FISs can be developed using either expert-driven or data-driven approaches (Awad and Fayek 2013). However, the curse of dimensionality results from the large number of model input variables in CLP models, as the number of rules grows exponentially with the number of input variables, and makes mapping the relationship between input variables and CLP complex, thus limiting the use of an expert-driven approach. To address this limitation of FISs in CLP modeling, a data-driven approach based on Fuzzy C-Means (FCM) clustering algorithm is used. FCM clustering is the most commonly used data-driven method for forming fuzzy sets and is commonly used for establishing the main patterns of the input–output data set (Pedrycz and Song 2012; Ahmad et al. 2012). Accordingly, the input–output data set of a given context, in the form $(\boldsymbol{x}_i, y), i = 1,2, \dots, N$, where input variables (factors, practices, and work sampling proportions) as $\boldsymbol{x}_i = [x_{1i}, x_{2i}, \dots, x_{ki}]$ and output (CLP) as $y = [y_i]$, is combined to form $(N+1)-$dimensional vector $\boldsymbol{p} = (\boldsymbol{x}_i, y)$. Then, using FCM clustering, a collection of C prototypes—$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_j\}$ and a partition matrix $U = [u_{ki}]$ representing the membership degree of a data instance in the j^{th} cluster—are developed by applying Eq. (5.1) in the product space of $\boldsymbol{X} \times Y$. The process results in C prototypes that each have a MF and corresponding to each of the fuzzy rules R_j, $j = 1,2, \dots, C$. Then, projecting the prototypes on the output space Y by considering their last coordinates as $v_1[y], v_2[y], \dots, v_j[y]$ results in the MFs of the output variable, which are denoted as $B_1, B_2, \dots, B_j$. Similarly, projecting the prototypes on the input space $\boldsymbol{X}$ as $\boldsymbol{v}_1[\boldsymbol{x}], \boldsymbol{v}_2[\boldsymbol{x}], \dots, \boldsymbol{v}_j[\boldsymbol{x}]$ results in the MFs of the input variables, which are denoted as $A_1, A_2, \dots, A_j$. This process results in a collection of rules of the form R_j : If X is A_j then y is B_j, $j = 1,2, \dots, C$. However, the projected MFs A_j and B_j do not come in a readily usable form as their shapes do not fit to common MFs shapes (e.g. triangular, trapezoidal, Gaussian, etc.) used in FIS models.

$$u_{ki} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|\boldsymbol{p_i} - \boldsymbol{v_i}\|}{\|\boldsymbol{p_i} - \boldsymbol{v_j}\|} \right)^{2/(m-1)}} \qquad (5.1)$$

The development of the rules and respective MFs was carried out using the MATLAB “genfis3 (Mamdani)” function. The function generated the respective fuzzy inference models using Fuzzy C-Means (FCM) clustering technique, and the projected input and output variable’s MFs A_j and B_j were approximated into a Gaussian membership functions. However, the Gaussian MFs are only approximations of the actual projected MFs, and this could reduce the accuracy of the developed FIS models. The fuzzification coefficient m in Eq. (5.1), commonly set at 2.0, indicates the degree of fuzziness of the developed MFs of the model variables. Lower m values (close to 1.0) will result in MFs that resemble the crisp set characteristic function of the model variables’ data, where membership values will be close to either 1 or 0; higher m values ($m = 3$) produce "spiky" MFs, where membership values are equal to 1 for data instances close to the prototypes, and decline in value the further they are from the prototypes (Pedrycz and Gomide 2007). The fuzzification coefficient m forms another model parameter for further optimization of fuzzy inference systems (Pedrycz et al. 2012). FCM clustering is carried out on the combined input–output space; the process results in a reduced number of rules, and counteracts the curse of dimensionality (Ahmad et al. 2012). However, using FCM clustering, only the rules that are representative of the data set are formed. Consequently, the total number of prototypes, C, determines the number of rules, which are equivalent to the number of MFs of each input and output variable, and forms another one of the key FIS model parameters for further optimization (Pedrycz et al. 2012).

The prediction process of FIS models involves coding inputs (i.e., fuzzifying input variables), input aggregation, fuzzy input-output implication, rule aggregation, and output decoding. In the coding inputs—or fuzzification—stage, the FIS model calculates the degree of membership for each fuzzy set based on the value of each input variable. Then, for each rule, a single value is calculated using input aggregation by applying a fuzzy operator to the membership values of each variable. Using a fuzzy operator, the different parts of the inputs in the conditions of the rules are combined, and outputs based on conclusions are implied using the if-then rule. The results of the different rules are then aggregated using a fuzzy operator and the output is predicted using an output decoding or defuzzification method. According to (Ahmad et al. 2012), FIS model accuracies are commonly determined using the RMSE. The RMSE of a CLP model is calculated using Eq. 5.2, where t_i is the target CLP value for the i^{th} data instance, y_i is the corresponding predicted CLP value, and n is the total number of data instances.

$$RMSE_l = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(t_i - y_i)^2} \tag{5.2}$$

An important interpretability measure of FIS developed using FCM clustering is the number of linguistic terms used to partition the model input and output variables and the number of rules (Ahmad et al. 2012). Maintaining a small number of linguistic terms and rules will make the rule base clearer for users to understand and therefore easier for them to interpret and put into practice (Gacto et al. 2011).

However, similar to the use of higher-order polynomial functions in regression analysis, FIS model accuracy will improve with a larger number of rules, as a larger number of rules will fit the data more accurately. On the other hand, since the number of rules is equivalent to the number of linguistic terms used for each model variable, having a large number of rules will result in very long if-then statements. For example, consider two FISs: one with two rules and the other with five. The first FIS model has two linguistic terms for each of the input and output variables; therefore, CLP, the output variable of the model, can only be linguistically represented by either of the terms “low” or “high”. The second FIS model has five linguistic terms for each model variable; therefore, the model’s CLP output can be linguistically expressed as “very low”, “low”, “medium”, “high”, or “very high”—a less interpretable result. In short, increasing the number of rules in an FIS model of CLP, though improving accuracy, would significantly reduce the interpretability of the model for users.

Additionally, a significant limitation of FISs is that they lack the ability to learn from data and optimize their model parameters. FIS-based models contain several parameters that can be optimized to improve the accuracy and interpretability of developed CLP models. Thus, there is a necessary trade-off between accuracy and interpretability in the optimization of FISs (Gacto et al. 2011). Therefore, in line with the interpretability requirement of FIS-based models, the maximum number of linguistic terms, and subsequently the number of rules, is limited to seven—the recommended number of items for accurate human interpretation (Pedrycz and Gomide 2007)—and the smallest number of rules, according to past studies with FCM clustering driven fuzzy inference systems, is three (Ahmad et al. 2012).

Furthermore, the interpretability of the fuzzy partitions is verified using completeness or coverage, normality, and distinguishability properties (Botta et al. 2009). The application of the FCM clustering approach ensures the coverage of the universe of discourse of each variable and the normality or full membership of the fuzzy sets at prototype values. The overlap of the fuzzy sets in the fuzzy partitions has to be kept to a level that the each couple of fuzzy sets are distinguishable enough (Botta 2008). The overlap between fuzzy sets is measured using a possibility measure Π (Eq. 5.3).

$$\Pi(A,B) = \sup_{x \varepsilon U} min\{\mu_A(x), \mu_B(x)\} \tag{5.3}$$

Possibility measure is recommended for efficient computation of distinguishability of fuzzy sets using overlap (Mencar et al. 2007), and to ensure interpretability of fuzzy systems, the overlap of the fuzzy sets shall not exceed 0.8 (Pulkkinen and Koivisto 2010).

5.4: Data Collection and Development of Contexts

The data for concreting (concrete placement) activity shown in Table 2.9 is used for investigating the role of context in CLP modeling. The concrete data was gathered from six building projects in the greater Edmonton area of Alberta, Canada. Data collection took place between June 2012 and October 2014 in collaboration with two partnering companies. The first three projects were built by Company 1, a multinational construction company with over 100 years of experience, and the last three were built by Company 2, a local construction company with over 40 years of experience. The projects included: (1) a commercial mixed-use office and staff facility building, (2) an industrial warehouse building, (3) a commercial warehouse building, (4) a mixed residential and community center building, (5) a high-rise mixed commercial-residential building, and (6) an institutional building. Concreting activity was studied in three data collection cycles, where each cycle extended over a month-long period and encompassed different weather seasons. For each data collection case, work sampling observations were made for the crew under study, and parameters (factors and practices), total work-hours, and installed quantities were documented.

This study's operational definition of context for CLP modeling, which states that context is what constrains the four elements of a CLP model (user, model developer, model, and prevailing environment of the model) without intervening in the model development process explicitly, was applied to the six projects. Using the answers to the 5W1H questions, the projects that shared similar contexts were grouped together, resulting in four unique contexts. The following context attributes distinguished the six projects that were studied, and, thus, were used as the key context attributes in comparing and identifying the similarity of the projects: "Who" attributes, related to the project owner's primary driver (schedule, cost, quality, or safety), contractor team's experience, and contractor organization's experience; "What" attributes related to project (i.e., building) type, site layout, project safety practice, and project productivity measurement and tracking practice; and "Where" attributes related to project location. Accordingly, projects having identical answers to key context attributes were grouped together, and the project (building type) context variable was used to name the four unique contexts.

Context 1, representing concreting in industrial buildings, includes the data sets of the first two projects (the commercial mixed-use office and staff facility building and the industrial warehouse building). The context attributes used to characterize the first project of Context 1 are shown in Table 5.1. Context 2, representing concreting in warehouse buildings, includes the data set of the third project (the commercial warehouse building). Context 3, representing concreting in high-rise buildings, includes the data sets of the fourth and fifth projects (the mixed residential and community center building and the high-rise mixed commercial-residential building). Context 4, representing concreting in institutional buildings, includes the data set of the sixth project (the institutional building). Additionally, the four context-specific data sets were combined to check if the context-specific models perform better than a generic model representing concreting in any building type and developed using the combined data set, and to verify the importance of applying context in CLP modeling.

5.5: ESTABLISHMENT OF CONTEXT-SPECIFIC MODEL FEATURES

The large feature space, made up of the 176 influencing parameters based on the moderated system approach, if unmodified would make the condition part of the if-then rules difficult to understand and interpret by model users; it must be reduced to improve the interpretability of the FIS. As discussed in Section 3.4, in this study, correlation-based feature selection (CFS) was applied for this purpose. The CFS algorithm was selected for its ability to deal with the high dimension of the features space and the small number of data instances, while preserving the original representation of the parameters and providing better understanding of the underlying process that generated the data (Guyon and Elisseeff 2003).

Using the Waikato Environment of Knowledge Analysis (WEKA) tool, the most relevant features or model variables having high correlations with CLP but low inter-correlations were identified for use in developing the context-specific models. A total of 16, 7, 8, and 11 features, representing the key influencing parameters, were selected for Context 1, Context 2, Context 3, and Context 4 CLP model developments, respectively. A similar feature selection process was also applied to the data set formed by combining the four context-specific data sets, and 16 features were established. The selected key influencing model features, together with their measurement scales, are shown in Table 5.2. The selected features clearly reveal that the key influencing features differed from one context to another. This finding is consistent with a previous study by Tsehayae and Fayek (2014), which compared the key influencing features, made up of key input parameters, among building and industrial contexts and showed that they are context-dependent and vary for the studied contexts.

5.6: Development and Performance of Original Context-Specific CLP Models

Based on the selected features shown in Table 5.2 and associated data sets of each context, the four context-specific fuzzy inference CLP models were developed in an original-case form, as illustrated in the flow chart shown in Fig. 5.1. Additionally a fifth generic CLP model based on the combined data set was developed in an original-case form. For all five original-case models, the following recommended model parameters were used: seven rules, a fuzzification coefficient $m = 2.0$, the MIN (minimum) fuzzy operator for input aggregation and implication, the MAX (maximum) operator for rule aggregation, and the CENTROID method for defuzzification (Pedrycz and Gomide 2007). Context 1, Context 2, Context 3, Context 4, and the generic CLP models had $RMSE_{OC}$ values of 1.582, 0.586, 2.193, 1.441, and 3.329, respectively. Comparisons indicated that all context-specific models performed better than the generic CLP model. The context-specific CLP models showed prediction accuracy superior to the generic CLP model by 52.48%, 82.41%, 34.12%, and 56.70%, respectively. These results indicate that a clear definition of context is useful to characterize and classify construction projects and assists in the process of grouping similar projects for more accurate CLP model development.

Table 5.2: Context-Specific CLP Modeling: Key Influencing Model Variables or Features

Feature, ID	Scale of measure	Context				
		1	2	3	4	G
Crew size $_{x1}$	Integer (total number of crew members)				•	
Craftsperson on-job training $_{x3}$	Real number (total duration of training, hours)	•				
Crew composition $_{x5}$	Proportion (ratio journeyman: apprentice)				•	
Co-operation among craftspersons $_{x8}$	1–5 predetermined rating				•	
Craftsperson motivation $_{x11}$	1–5 predetermined rating					•
Craftsperson fatigue $_{x12}$	Real number (ratio total weekly worked hours: regular weekly work hours)					•
Craftsperson trust in foreman $_{x13}$	1–5 predetermined rating					•
Team spirit of crew $_{x14}$	1–5 predetermined rating	•				
Level of absenteeism $_{x15}$	Real number (ratio absent crew member: total crew size)					•
Level of interruption and disruption $_{x18}$	1–5 predetermined rating				•	
Fairness of work assignment $_{x19}$	1 - 5 predetermined rating				•	
Job security $_{x23}$	Real number (average length of unemployment period, months)					•
Availability of tools $_{x37}$	Real number (average waiting time, minutes)					•
Concrete placement technique $_{x45}$	Categorical: pump (1), crane and bucket (2), direct chute (3)	•	•	•		•
Structural element type $_{x46}$	Categorical: columns (1), footings (2), grade beams (3), pile caps (4), slabs (5), walls (6)	•	•	•	•	•
Location of work scope, distance $_{x48}$	Real number (distance, m)		•			
Location of work scope, elevation $_{x49}$	Real number (elevation, m)	•				
Congestion of work area $_{x50}$	Real number (ratio of peak to average manpower)		•			
Fairness in performance review of crew by foreman $_{x57}$	1–5 predetermined rating	•		•		
Change of foremen $_{x58}$	Integer (number of turnovers per month)					•
Site congestion $_{x90}$	Real number (ratio free site space: total site area)				•	
Project work times $_{x96}$	Real number (ratio total worked hours per day)					•

Table 5.2: Context-Specific CLP Modeling: Key Influencing Model Features (continued)

Feature, ID	**Scale of measure**	**Context**				
		1	**2**	**3**	**4**	**G**
Owner staff on site x_{97}	Integer (total number of owner staff on site)					•
Approval of shop drawings and sample materials x_{101}	Real number (average time taken to approve, days)					•
Treatment of foremen by superintendent and project manager x_{106}	1–5 predetermined rating	•				
Uniformity of work rules by superintendent x_{108}	1–5 predetermined rating	•	•			
Labour productivity measurement practice x_{126}	1–5 predetermined rating					•
Out-of-sequence inspection x_{129}	Real number (number of occurrences per week)	•				
Safety training x_{135}	Real number (total duration of training, hours)	•				
Project safety administration and reporting x_{138}	1–5 predetermined rating	•				
Availability of communication devices x_{143}	Real number (ratio number communication radios: total number of crews)					•
Oil price fluctuation x_{167}	Real number (weekly price change, %)	•				
Natural gas price x_{168}	Real number (dollar / GJ)	•				
Direct work proportion y_1	Proportion (%)	•	•	•	•	•
Preparatory work proportion y_2	Proportion (%)	•		•		
Tools and equipment proportion y_3	Proportion (%)	•		•	•	
Material handling proportion y_4	Proportion (%)			•	•	
Travel proportion y_6	Proportion (%)			•	•	
Personal proportion y_7	Proportion (%)		•			•

Note: G context represents the generic CLP context, based on combined data set. There were 23, 16, 28, 25, and 92 data instances for contexts 1, 2, 3, 4, and generic, respectively.

5.7: OPTIMIZATION AND VALIDATION OF CONTEXT-SPECIFIC CLP MODELS

FIS models contain several model parameters that can be optimized to improve the accuracy and interpretability of the developed CLP models. The model parameters available for optimizing FIS-based models include: the fuzzification coefficient (m) in the FCM clustering; MF parameters; number of rules; confidence levels of rules (degree of support of rules); fuzzy operators for input aggregation, implication, and rule aggregation; and defuzzification methods. Optimizing these numerous model parameters at once will create a large search space, so optimizing each parameter separately is actually more efficient. Doing so will create a smaller search space where the optimization process has a better chance of arriving at optimal model parameter values (Cordón 2011).

Thus, in this study, the context-specific models were optimized by dividing the search space into smaller spaces and conducting a stage-by-stage optimization of: (1) the fuzzification coefficient m in FCM clustering, (2) MF parameters, (3) number of rules, and (4) fuzzy operators and defuzzification methods. Fig. 5.1 describes the optimization steps. The rules' confidence levels or degrees of support were adjusted, as the rules were generated using FCM clustering, which generates unique and equally representative rules for the data set. Among the four model parameters optimized; the number of rules—equivalent to the total number of prototypes C developed in FCM clustering process, which in turn is equivalent to the number of linguistic terms of each model variable—is the most critical component of the optimization process. Thus, based on the interpretability requirement, the number of rules was increased incrementally from a minimum of three to a maximum of seven, creating five interpretable case models. The optimization process was carried out for each of the context-specific models and also for the generic CLP model, and was implemented in the MATLAB® 2014 environment.

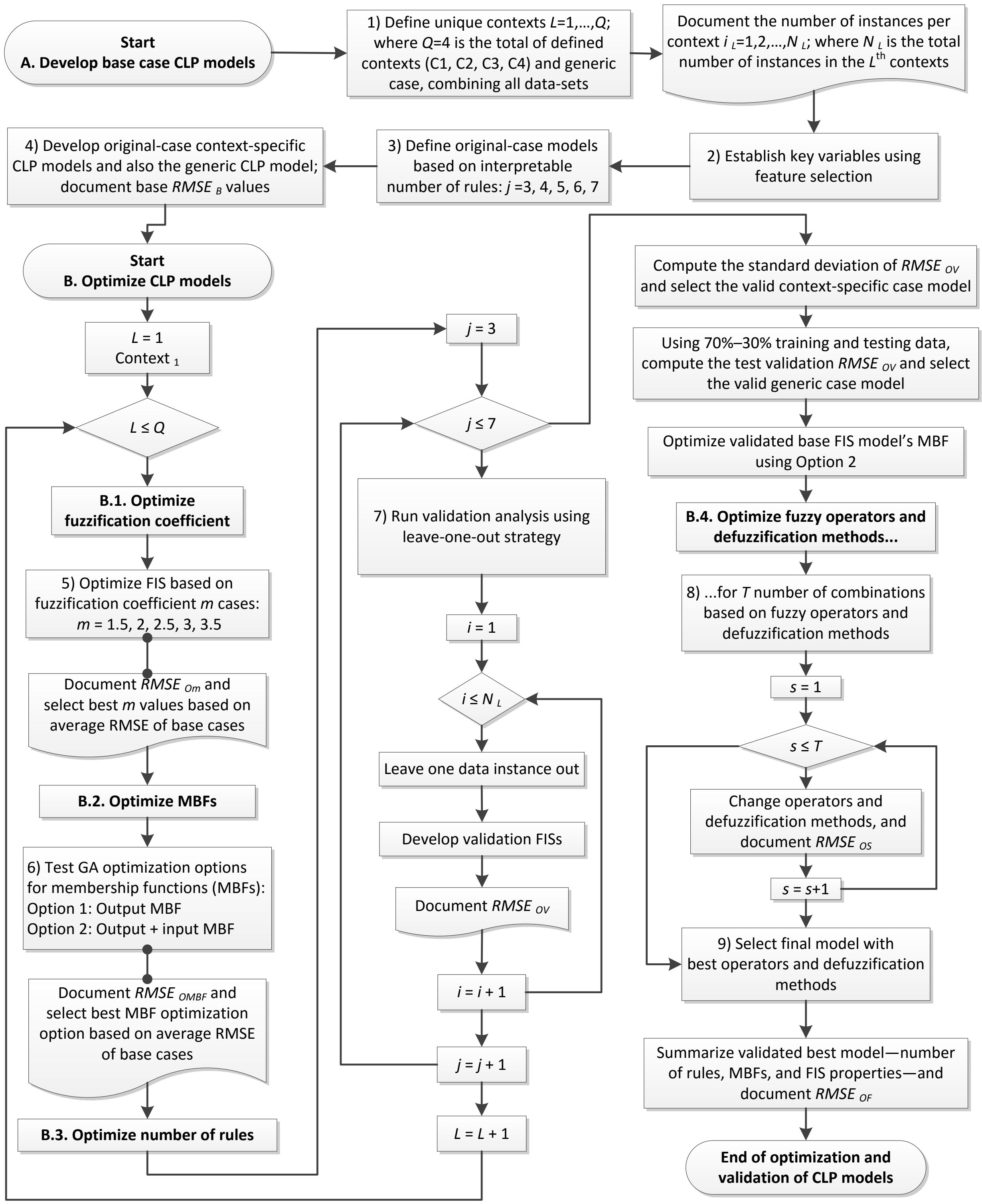

Figure 5.1: Flow Chart for Context-Specific CLP Model Development, Optimization, and Validation

5.7.1: Optimization of Fuzzification Coefficient

The fuzzification coefficient m in the FCM clustering is used to develop the rule bases and MFs of the context-specific CLP models. Typically, the value of m is assumed to be equal to 2.0, although values between 1.5 and 3.5 have been examined in the optimization of FCM-driven FIS models (Pedrycz et al. 2012). Since the fuzzification coefficient m directly affects the shape of the MFs, which are equivalent to the number of FIS model rules, the fuzzification coefficient m was optimized for each of the case models. The best-performing model parameters were then established by averaging the $RMSE_{Om}$ values of each case model and taking the one resulting in the lowest average $RMSE_{Om}$ value. Accordingly, five fuzzification coefficient cases where $m = 1.5$, $m = 2.0$, $m = 2.5$, $m = 3.0$, and $m = 3.5$ were investigated for the four context-specific and the generic CLP models. The results of the optimization process are shown in Table 5.3. The optimum m values for the Context 1, Context 2, Context 3, Context 4, and generic CLP models were equal to 1.5, 2.5, 2.0, 2.0, and 1.5, respectively.

Table 5.3: CLP Models Optimization for Fuzzification Coefficient m

Context	**Average RMSE [a]**				
	m = 1.5	**m = 2.0**	**m = 2.5**	**m = 3.0**	**m = 3.5**
1	**1.699**[b]	1.712	1.961	2.012	2.012
2	0.705	0.556	**0.491**[b]	0.514	0.505
3	6.005	**2.021**[b]	2.273	3.139	3.310
4	1.970	**1.622**[b]	1.939	2.037	2.164
Generic	**2.657**[b]	3.054	3.038	3.305	3.514

[a] RMSE values are based on the average RMSE values of the five case models
[b] Denotes the lowest average RMSE value

5.7.2: Optimization of Membership Function Parameters

MFs define the degree to which an element of a model feature belongs to any one of the linguistic terms characterizing the feature, and they are highly context-specific (Awad and Fayek 2013). In this study, Gaussian MFs based on Eq. (5.4) have been used for both the input features, made up of key influencing CLP variables, and the output, or CLP, feature and their associated linguistic terms. For a given feature with universe of discourse of x and linguistic term A, the MF parameter σ represents the standard deviation, denoting the spread of A, and μ represents the modal value, denoting the typical element of A; both parameters were optimized to improve the accuracy of the developed CLP models.

$$A(x, \sigma, \mu) = e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{5.4}$$

The optimization is carried out over the MF parameters $[\sigma, \mu]$ of each of the CLP model's features, resulting in a nonlinear large search space that requires an evolutionary optimization technique. GA is the most commonly used evolutionary optimization technique for finding optimal or near-optimal solutions for a given search space and have been successfully applied in the tuning or optimization of membership functions in fuzzy inference models, as the objective function for optimizing such problems is complex and nonlinear, and thus cannot be optimized using traditional gradient based techniques (Cordón 2011). For each of the context-specific and the generic models, two MF optimization options were considered. Considering P input features, one output feature, and C rules, the first option (Option 1) was based on optimizing all of the MFs in the input and output space simultaneously, resulting in search space U_1, where $U_1 = 2 * (P + 1) * C$. The second option (Option 2) was a subset of Option 1 and was based on optimizing only the output, or CLP feature, MFs.

The optimization process using GA required encoding schemes to transform the MFs of the context-specific model features into a chromosome. The real chromosome coding structure was formulated based on the two parameters $[\sigma, \mu]$ of each linguistic term of the model features. Then, the genetic operations of reproduction, crossover, and mutation were performed. Each operation generated new sets of chromosomes, representing a new solution that meets the optimization constraints. The solution chromosomes were checked according to the following MF optimization constraints: (1) The standard deviation σ_j representing the spread of any linguistic term must be greater than zero; (2) The modal value μ_1 of the first linguistic term of any model feature must not be less than the minimum value of the feature; (3) The modal value μ_C of the last linguistic term of any model feature must not be greater than the maximum value of the feature; (4) The modal value μ_j of the j^{th} linguistic term must not be greater than that of the modal value μ_{j+1} of the next linguistic term, and (5) The overlap between j^{th} and $j+1^{th}$ linguistic terms shall not exceed the limit value of 0.8.

The optimization process was implemented in MATLAB® 2014 environment, and started with 150 randomly generated initial population of solutions, and used expert judgment to identify the possible solutions which meet the above MF optimization constraints. Expert judgment was used to evaluate the initial random solutions as past solutions for the highly complex and nonlinear CLP problem did not exist. The objective function of the genetic search was to minimize the $RMSE_{OMF}$ of each CLP model; the fitness value of each solution was determined by calculating the $RMSE_{OMF}$, which was then used as a parent for the development of the next solution using crossover and mutation. The crossover swaps parts of two chromosomes (i.e., solutions) according to a crossover probability, and creates the next chromosomes. Then after, if the average fitness of the new solution was smaller than the average fitness of the previous solutions, a random change in the information of the new solution was carried out according to the mutation probability. Crossover probability of 0.8 and mutation probability of 0.01 was used, and the stopping criterion was based on the fitness limit, where the iteration stopped when the fitness value of the last solution is greater than the best fitness value. Finally, the overlap among the MFs in the best solution was evaluated using the degree of overlap.

However, due to the limited data available for the optimization process, the interpretability constraint limiting the degree of overlap among optimized MFs to 0.8 was not always fulfilled, and additional expert visual assessment of the MFs plot was carried out to verify the interpretability of the optimized MFs. Accordingly, in cases when overlap among successive MFs exceeded the limit value of 0.8 and the visual assessment verified lack of interpretability of MFs, the developed solution is rejected and the optimization process is repeated with new randomly generated solutions.

Similar to the optimization process adopted for the fuzzification coefficient, the MF optimization process was investigated for each of the five case models of a given context. For each CLP model, the two MF optimization options were tested and the $RMSE_{OMF}$ value of each optimized case model was calculated. Then, the average $RMSE_{OMF}$ of the five submodels was calculated for each option and compared against the $RMSE_{OMF}$ value of the original-case model. The results of the MF optimization process, shown in Table 5.4, indicate that Option 1, which optimized all of the MFs in the input and output space simultaneously, did not show an improvement in $RMSE_{OMF}$ values. This option was not effective, as the search space was too large, and the associated data instances were small and unable to improve model accuracy. Option 2, which optimized only the output feature MFs, showed 0.58%, 5.71%, 23.00%, 15.91%, and 16.61% improvements of average $RMSE_{OMF}$ over the original-case CLP models for Context 1, Context 2, Context 3, Context 4, and the generic case, respectively.

Table 5.4: CLP Models Membership Functions Optimization

Context	**Average RMSE [a]**			
	Before optimization	**After optimization**		**Improvement (Option 2 over original case)**
	Original Case	**Option 1**	**Option 2**	**%**
1	1.582	2.467	1.573	0.58
2	0.586	0.865	0.552	5.71
3	2.193	7.392	1.689	23.00
4	1.441	3.172	1.212	15.91
Generic	3.329	5.150	2.776	16.61

[a] RMSE values are based on the average RMSE values of the five case models

5.7.3: Optimization of Number of Rules and Validation of CLP Models

The quality of the developed context-specific CLP models is guided by two main criteria: accuracy and interpretability. The interpretability of FIS models developed using FCM clustering is directly related to the number of rules (Cordón 2011). Accordingly, five interpretable case models with the number of rules ranging from three to seven were developed for each context. Furthermore, in fuzzy inference system modeling it is important to confirm that the developed fuzzy models are properly validated so as to ensure they are capable of producing meaningful results. As the data sets in context-specific CLP modeling are often small, the use of a leave-one-out validation strategy is recommended (Pedrycz and Gomide 2007), as splitting the data into, for example, 70%–30%, will result in a very small number of data points available for model development. Thus, using this strategy for each of the five case models of a context-specific model, all but one data point is used in model development and then the model performance, in terms of $RMSE_{OV}$, is evaluated using the left-out data instance. This process is repeated for all data instances, and N numbers of error measures based on $RMSE_{OV}$ values are determined, where N represents the number of data instances in a given context (refer to Fig.1). The lowest standard deviation value of the $RMSE_{OV}$ is then used to evaluate and identify the number of rules leading to the most accurate and valid context-specific CLP model.

For the fifth generic CLP model, which is based on a relatively large combined data set, a traditional 70%–30% split of the data into training and testing subsets was used to validate the five case models (Awad and Fayek 2013). Thus, 70% of the data set was randomly selected and used for the development of the five cases of the generic models, with the number of rules ranging from three to seven, and the remaining 30% of the data set was used to test the developed models and establish the $RMSE_{OV}$ for validation. The lowest $RMSE_{OV}$ value, based on the test data, was used to identify the most valid generic case model. For each of the context-specific and the generic CLP models, the optimized fuzzification coefficient values were used in the validation process. The results of the validation process are shown in Table 5.5.

Table 5.5: CLP Model Validation and Optimization of Number of Rules

Context	RMSE									
	No. of rules = 3		No. of rules = 4		No. of rules = 5		No. of rules = 6		No. of rules = 7	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
1	1.488	1.389	1.567	1.312	1.502	1.272	1.453	**1.255**[a]	1.348	1.315
2	0.679	0.543	0.655	0.530	0.574	0.521	0.610	0.508	0.647	**0.502**[a]
3	2.492	3.131	2.348	3.072	2.380	2.907	2.320	**2.845**[a]	2.297	3.145
4	2.078	1.328	2.003	1.409	2.022	1.305	1.974	1.301	1.985	**1.270**[a]
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Generic	3.198	3.016	2.944	**1.885**[b]	2.907	1.901	2.938	2.806	2.929	2.658

Note: [a] Denotes the lowest RMSE standard deviation value of a context-specific model and [b] denotes the lowest testing RMSE value of the generic model, indicating the most valid models.

Accordingly, the numbers of rules leading to the most valid context-specific model are equal to 6, 7, 6, and 7 for the Context 1, Context 2, Context 3, and Context 4 CLP models, respectively. For the generic CLP model, the case model having the lowest validation $RMSE_{OV}$ had 4 rules. The valid context-specific and generic CLP models were highly interpretable and stable against variations in the experimental data used for model development. Each valid model's output or CLP membership function parameters were optimized using the second MF optimization option and the overlap among the developed membership functions of each model were constrained using the possibility measure and the interpretability of the fuzzy sets was verified.

5.7.4: Optimization of Fuzzy Operators and Defuzzification Methods

In fuzzy inference systems, not only the MFs but also the fuzzy operators and defuzzification methods are context-dependent (Klir and Yuan 1995). Based on the results of the optimum model parameters for the fuzzification coefficient, optimized output (CLP) membership functions using GA, and numbers of rules, as shown in Table 5.6, the optimized CLP models were developed and sensitivity analysis was carried out for the fuzzy operators and defuzzification methods.

Table 5.6: Optimized CLP Models: Structure and Model Parameters

FIS structure and model parameters	**Context**				
	1	**2**	**3**	**4**	**Generic**
	Concreting, industrial buildings	**Concreting, warehouse buildings**	**Concreting, high-rise buildings**	**Concreting, institutional buildings**	**Concreting, any building**
Number of input features	16	7	8	12	16
Fuzzification coefficient	1.5	2.5	2.0	2.0	1.5
Number of rules	6	7	6	7	4
Input aggregation operator	PROD	MIN	PROD	PROD	PROD
Implication method	PROD	MIN	PROD	PROD	PROD
Rule aggregation operator	MAX	SUM	PROBOR	PROBOR	MAX
Defuzzification method	MOM	BISECTOR	CENTROID	BISECTOR	CENTROID
Accuracy (RMSE)	1.162	0.467	0.992	0.671	2.515

The following options of fuzzy operators and defuzzification methods were tested: for input aggregation [MIN (minimum) and PROD (product)], for implication [MIN (minimum) and PROD (product)], for rule aggregation [MAX (maximum), SUM (sum of each rule's output set), and PROBOR (probabilistic OR)], and for defuzzification [CENTROID, BISECTOR, MOM (middle of maximum), LOM (largest of maximum), and SOM (smallest of maximum)]. The options were varied one at a time, and a total of 30 unique combinations were tested. The options and results for Context 4 CLP model are shown in Table 5.7. For each optimized CLP model, the $RMSE_{OS}$ was determined based on the best combination of the listed options of fuzzy operators and defuzzification methods; the results yielding the lowest $RMSE_{OS}$ values provided the best performing fuzzy operators and defuzzification methods for CLP models, as shown in Table 5.6.

The final optimized context-specific and generic CLP models parameters are as shown in Table 5.6. The final optimized Context 1, Context 2, Context 3, Context 4, and generic CLP models had $RMSE_{OF}$ values of 1.162, 0.467, 0.992, 0.671, and 2.515, respectively. Comparing the final optimized models against the original-case models indicated that the optimization process improved the accuracy of the Context 1, Context 2, Context 3, Context 4, and generic CLP models by 26.52%, 20.22%, 54.78%, 53.44%, and 24.46%, respectively. Furthermore, comparing the final optimized CLP models with each other indicated that all context-specific models performed better than the generic CLP model.

In summary, applying the operational definition of context to CLP modeling resulted in the development of unique contexts derived from studied construction projects. Notably, the context-specific CLP models addressing the unique contexts resulted in more accurate predictions than the generic model did; this finding supports the need for careful examination of context in CLP research (Green et al. 2009). The models developed through this study can be used to predict the CLP of concreting activities for new projects, either using context-specific CLP models in cases where a given new project's context attributes based on 5W1H questions resemble any one of the studied contexts, or using the generic CLP model.

Table 5.7: CLP Models Optimization of Fuzzy Operators and Defuzzification Methods

Case	Fuzzy operators and Defuzzification methods				
	Input aggregation	**Implication method**	**Rule aggregation**	**Defuzzification method**	**Accuracy (RMSE)**
1	MIN	MIN	MAX	CENTROID	0.926
2	MIN	MIN	MAX	BISECTOR	0.904
3	MIN	MIN	MAX	MOM	0.845
4	MIN	MIN	MAX	LOM	1.441
5	MIN	MIN	MAX	SOM	1.360
6	MIN	MIN	SUM	CENTROID	0.908
7	MIN	MIN	SUM	BISECTOR	0.876
8	MIN	MIN	SUM	MOM	0.820
9	MIN	MIN	SUM	LOM	1.441
10	MIN	MIN	SUM	SOM	1.324
11	MIN	MIN	PROBOR	CENTROID	0.908
12	MIN	MIN	PROBOR	BISECTOR	0.876
13	MIN	MIN	PROBOR	MOM	0.820
14	MIN	MIN	PROBOR	LOM	1.441
15	MIN	MIN	PROBOR	SOM	1.324
16	PROD	PROD	MAX	CENTROID	0.674
17	PROD	PROD	MAX	BISECTOR	0.683
18	PROD	PROD	MAX	MOM	0.703
19	PROD	PROD	MAX	LOM	1.441
20	PROD	PROD	MAX	SOM	0.703
21	PROD	PROD	SUM	CENTROID	0.674
22	PROD	PROD	SUM	BISECTOR	0.683
23	PROD	PROD	SUM	MOM	0.683
24	PROD	PROD	SUM	LOM	1.441
25	PROD	PROD	SUM	SOM	0.703
26	PROD	PROD	PROBOR	CENTROID	0.674
27	PROD	PROD	PROBOR	BISECTOR	**0.671**
28	PROD	PROD	PROBOR	MOM	0.692
29	PROD	PROD	PROBOR	LOM	1.441
30	PROD	PROD	PROBOR	SOM	0.692

5.8: Chapter Summary

CLP is a function of various controllable and uncontrollable influencing variables; to improve CLP, the effects of these variables must be identified. This chapter studied the effects of numerous objective and subjective variables made up of factors, practices, and work sampling proportions. It also suggested an operational definition of context and developed associated context attributes to explicitly define the context of a given project under investigation. Based on the context attributes, the six projects studied were grouped into four unique contexts for which four context-specific CLP models were developed. Using data-driven feature selection techniques, the key influencing variables, or features, were established, and comparing the key features uncovered significant difference among contexts.

Subsequently, using data-driven fuzzy inference systems, four context-specific models and a generic CLP model (developed by combining the four context-specific data sets) were developed. All four context-specific models showed superior prediction accuracy when compared to the generic model. This study further examined the use of hybrid fuzzy techniques to overcome the inability of FIS to learn from data. In this study, the following FIS model parameters were optimized: (1) the fuzzification coefficient m in FCM clustering, (2) membership function parameters, (3) number of rules, and (4) fuzzy operators and defuzzification methods. The optimization process improved the accuracy and interpretability of the developed context-specific and generic CLP models. However, the FIS model development and optimization process has some limitations. The actual MFs projected from the FCM clusters were approximated using Gaussian membership functions, and this could reduce the accuracy of the developed FIS models. Thus, exploring other approximations of the projected MFs using triangular or trapezoidal shapes is recommended so as to investigate the effect of the approximated Gaussian MFs on the accuracy of the developed models. Additionally, during the membership function optimization process, the initial solutions were evaluated using expert judgment, also the overlaps among MFs of the final best solution were evaluated using expert visual assessment. Thus, collecting additional data to establish solutions and further expand the number of data instances for optimization, and introducing additional constraints on the standard deviation parameter of the MFs to improve the interpretability of optimized

MFs are recommended so as to improve the MF optimization process. In the next chapter, further research on context adaptation and its application in CLP modeling by developing an approach that allows for the adjustment of the context-specific model's membership functions from one context to another, thereby transferring CLP knowledge bases represented in the context-specific CLP models among contexts is presented.

5.9: References

Ahmad, S. S. S., Ahmad, S., and Pedrycz, W. (2012). “Data and feature reduction in fuzzy modeling through particle swarm optimization.” *Appl. Comput. Intell. Soft Comput.,* 2012, 21 pp.

Awad, A., and Fayek, A. R. (2013). "Adaptive learning of contractor default prediction model for surety bonding." *J. Constr. Eng. Manage*., 139(6), 694–704.

Bazire, M., and Brézillon, P. (2005). “Understanding context before using it.” *Modeling and using context,* Springer, New York, US, 29–40.

Botta, A. (2008). “Automatic context adaptation of fuzzy systems.” Ph.D. thesis, IMT Institute for Advanced Studies, Lucca, Italy.

Brézillon, P., and Pomerol, J. (2001). “Modeling and using context for system development: Lessons from experience.” *J. Decis. Systems,* 10(2), 265–288.

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.

Christian, J., and Hachey, D. (1995). "Effects of delay times on production rates in construction." *J. Constr. Eng. Manage*., 121(1), 20–26.

CII (Construction Industry Institute). (2010). *Guide to activity analysis*, Univ. of Texas at Austin, Austin, Texas, US.

Cordón, O. (2011). “A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems.” *Intern. J. Approx. Reasoning,* 52(6), 894–913.

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

Fayek, A. Robinson, and Tsehayae, A. A. (2012). "Modeling construction labour productivity using fuzzy logic and exploring the use of fuzzy hybrid techniques." *Fuzzy Information Processing Society (NAFIPS), 2012 Annual Meeting of the North American,* IEEE, Berkeley, California, US, 1–6.

Gacto, M. J., Alcalá, R., and Herrera, F. (2011). “Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures.” *Inf. Sci.,* 181(20), 4340–4360.

Gerek, İ. H., Erdis, E., Mistikoglu, G., and Usmen, M. (2014). “Modelling masonry crew productivity using two artificial neural network techniques.” *J. Civ. Eng. Manage.,* (ahead-of-print), 8pp.

Green, S. D., Kao, C., and Larsen, G. D. (2009). “Contextualist research: Iterating between methods while following an empirically grounded approach.” *J. Constr. Eng. Manage.,* 136(1), 117–126.

Guyon, I., and Elisseeff, A. (2003). “An introduction to variable and feature selection.” *J. Machine Learn. Res.,* 3, 1157–1182.

Jang, S., and Woo, W. (2003). “Ubi-UCAM: A unified context-aware application model.” *Modeling and Using Context,* Springer, New York, US, 178–189.

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

Mao, H. (1999). “Estimating labour productivity using fuzzy set theory.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Mencar, C., Castellano, G., and Fanelli A.M. (2007). "Distinguishability quantification of fuzzy sets." *Inform. Sci.*, 177, 130–149.

Oral, M., Oral, E. L., and Aydın, A. (2012). “Supervised vs. unsupervised learning for construction crew productivity prediction.” *Autom. Constr.*, 22, 271–276.

Pedrycz, W., and Gomide, F. (2007). *Fuzzy systems engineering: Toward human-centric computing.* John Wiley and Sons, Hoboken, New Jersey, US.

Pedrycz, W., Russo, B., and Succi, G. (2012). “Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity.” *App. Soft Comput.,* 12(8), 1985–1995.

Pedrycz, W., and Song, M. (2012). “Granular fuzzy models: A study in knowledge management in fuzzy modeling.” *Intern. J. Approx. Reasoning,* 53(7), 1061–1079.

Pulkkinen, P., and Koivisto, H. (2010). "A dynamically constrained multiobjective genetic fuzzy system for regression problems." *IEEE T. Fuzzy Syst.,* 18, 161–177.

Shahtaheri, M., Nasir, H., and Haas, C. T. (2014). "Setting baseline rates for on-site work categories in the construction industry." *J. Constr. Eng. Manage.,* 141(5), 04014097.

Song, L., and AbouRizk, S. M. (2008). "Measuring and modeling labor productivity using historical data." *J. Constr. Eng. Manage*., 134(10), 786–794.

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). "Modeling construction labor productivity." *J. Constr. Eng. Manage*., 116(4), 705–726.

Thomas, H. R., and Raynar, K. A. (1997). "Scheduled overtime and labor productivity: quantitative analysis." *J. Constr. Eng. Manage*., 123(2), 181–188.

Thomas, H., and Yiakoumis, I. (1987). “Factor model of construction productivity.” *J. Constr. Eng. Manage.,* 113(4), 623–639.

Tsehayae, A. A., and Fayek, A. Robinson. (2014). “Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects.” *Can. J. Civ. Eng.,* 41(10), 878–891.

Turner, R. M. (2014). “Context-mediated behavior.” *Context in computing,* Springer, New York, US, 523–539.

Yi, W., and Chan, A. (2014). "Critical Review of Labor Productivity Research in Construction Journals." *J. Manage. Eng.,* 30(2), 214–225.

Zhu, Y. (2005). "Applying context to systems integration in AEC." *Proc., Constr. Research Congress*, ASCE, San Diego, California, US, 1–8.

Chapter 6: Context Adaptation of CLP Models[6]

6.1: INTRODUCTION AND BACKGROUND

Construction labour productivity (CLP) is one of the most studied areas in construction engineering and management field (Yi and Chan 2014). As a result, numerous predictive CLP models as shown in Fig. 2.1 have been developed. However, in CLP modeling field, an approach for transferring the knowledge represented in such models from one context to another context is missing. Such an approach is useful in modeling new contexts for which data availability is limited; and existing models cannot be applied without some adaptation (Ji et al. 2012). Model adaptation is of a particular importance for Mamdani-type fuzzy inference system based models, as the fuzzy sets used in such models are highly context-specific. Mamdani-type FISs are based on fuzzy if-then rules that relate the input parameters to the output parameter, where both inputs and outputs are represented using linguistic variables. Fuzzy sets are used to describe the respective linguistic variables, which characterize say a CLP influencing parameter like *crew size* using linguistics variables such as *small crew size*, *medium crew size*, and *large crew size* (refer to Fig. 6.1).

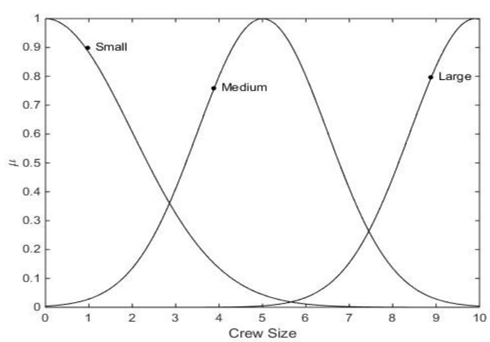

Figure 6.1: Gaussian Membership Functions for Fuzzy Sets Representing "*Crew Size*"

A fuzzy set A, representing a linguistic variable is characterized using its membership function, which represents numerically the degree to which an element x belongs to the fuzzy set and fits the linguistic variable over a continuous range $A\colon \boldsymbol{X} \rightarrow \mu = [0, 1]$ as shown in Fig. 6.1. In the fuzzy inference based CLP models, the input (key factors, practices, and work sampling proportions) and output (CLP)

---

6 Parts of this chapter have been submitted for publication in Journal of Construction Innovation: Information, Process, Management, JCI, 36 manuscript pages, submitted July 28, 2015.

feature were partitioned using linguistic variables over their respective universe of discourse. The linguistic variables were represented using Gaussian membership functions (refer to Fig. 6.1). The membership function for a Gaussian membership function $A(x, \sigma, \mu)$ is shown in Eq. (6.1), where σ represents the standard deviation, denoting the spread of A, and μ represents the modal value, denoting the typical element of A:

$$A(x, \sigma, \mu) = e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{6.1}$$

Membership functions (MFs) have many important descriptors. Membership functions employed in modeling endeavours are required to be normal, that is at least one element of $\boldsymbol{X}$ attains full membership ($\mu = 1$) and represents a typical value of the fuzzy set. Support of fuzzy set A represents all elements of $\boldsymbol{X}$ that exhibit some association with the fuzzy set by having nonzero membership degrees and core of fuzzy set A represents all elements of the universe $\boldsymbol{X}$ that are typical to A. In FIS, the if-then rules are composed of fuzzy conditions (represented by the membership functions of the input features) and fuzzy conclusions (represented by the membership functions of the output feature), an example of a fuzzy rule is shown below, where the words in italics are the features and the linguistic variables are shown in bold:

If the *crew size* is **medium** and the *crew composition* is **good** and the *co-operation among craftsperson* is **very good** and the *level of interruption and disruption* is **low** and the *fairness of work assignment* is **good** and the *site congestion* is **low** and the *structural element type* is grade beams and the *direct work proportion* is **high** and the *tools and equipment proportion* is **average** and the *material handling proportion* is **average** and the *travel proportion* is ***low*** then *construction labour productivity* is **high**.

Fuzzy inference systems consist of two main parts: a rule-base (RB) composed of the linguistic if-then rules and a database (DB) which associates membership functions to the linguistic variables used in the RB, e.g., **small** for crew size or **high** for CLP. However, it is worth nothing that in the database (DB) parts of FISs, the concept of "*small crew size*" cannot be uniquely and universally defined, while concepts like “crew size under 5” can be. It is natural that construction professionals could define such fuzzy sets differently depending on the context of use, and the context itself is defined using context attributes like type of projects, e.g. commercial or industrial. Thus, the exact definition of such fuzzy sets depends on an external parameter like type of project (Cordón et al. 2001), as for example, industrial projects tend to have larger crew sizes as compared to commercial ones. On the other hand, the rule-base is assumed to be context-free (Pedrycz et al. 1997), as it is usually valid, independent of the context of use. This is because the context of use affects only the database or the meaning associated with each linguistic term used in the rules rather than the logic of the rules themselves (Botta et al. 2009).

As discussed in Chapter 5, the main idea behind defining context in CLP studies is the idea of restriction, as when context is fixed, the universe of the system under study will also be restricted to a particular universe (Gudwin and Gomide 1994). Working on such restricted universes will modify the perception and the meaning of fuzzy sets used in such systems, indicating that fuzzy sets and in process the developed fuzzy inference systems are strongly context-dependent (Pedrycz and Gomide 2007). Therefore, the application of the developed fuzzy inference systems in new contexts will require some adaptation, as no past context is ever exactly the same as a new one, and old or base models should be adjusted to fit the new context (Ji et al. 2012).

Furthermore, context adaptation of fuzzy inference systems have been recognized as an effective method for generating interpretable and accurate FIS models by tuning the parameters contained in the database and using context-specific information (Botta et al. 2009). However, in the construction engineering and management research area, context adaptation of FIS is scantly explored and studies have rarely studied context adaptation in a formal way. In context adaptation, the objective is to use the model features (key influencing variables) and associated model of one context to model another one, after adaptation, while using the same model features of the first context. Thus, context adaptation will enable the reuse of existing models and knowledge bases in new contexts. This will save model developers the considerable effort required to collect data for all influencing variables, and to develop and optimize new models. It will also improve the implementation of existing models by industry as the existing models can be adapted to suit the industry's specific context or need.

In this chapter, a literature review of the approaches used for context adaptation of fuzzy inference systems is presented and viable options are identified. Next, the chapter discusses the procedure for adapting the series of context-specific CLP models discussed in Section 5.6 to suit new contexts based on the field data collected for the key variables or features of the respective CLP models. Subsequently, it performs a comparative assessment of the adapted and base context-specific models and discusses the effectiveness of context adaptation in CLP modeling. Finally, conclusions are presented.

6.2: Literature Review

Context plays an active part in construction research analysis as it is invariably dynamic and imperative for the development of meaningful findings (Engwall 2003), and subsequently, context adaptation has an important application in deriving new models from existing ones. However, in construction research field, context adaptation is not properly explored and studies have rarely studied fuzzy inference systems context adaptation in a formal way, but rather focused on the fine tuning of model parameters to improve model accuracy. The adaptation of a fuzzy inference model to improve the accuracy of predicting contractor's default was investigated, based on the tuning of membership functions and weight of rules of the FIS and using neural network and genetic algorithm techniques, and accuracy of the model was improved (Awad and Fayek 2013). Oduba (2002) also explored the tuning or calibration of the output membership functions of FIS based CLP models for industrial pipe rigging and welding activities, and relied on shifting the right, left, and both legs of the membership functions, and reported some improvement in models accuracy.

In the general computing field, most of the context adaptation research has been carried out on Mamdani-type fuzzy inference systems (FISs), which have been used in a wide range of areas due to their ability to handle linguist concepts and perform an accurate modeling of input-output relations (Botta

2008). In the context adaptation of FISs, the focus has been mainly on the linguistic variables and their respective membership functions as the rule-base is considered to be a context-free model (Gudwin et al. 1998; Gudwin and Gomide 1994). According to Botta et al. (2009), the following principles have been followed in adapting fuzzy inference systems: (1) context adaptation will not modify the rule-base as the rule-base is considered to be a context-free and universal knowledge; (2) context adaptation will not change the number of linguistic variables and, consequently, the number of corresponding fuzzy sets defined in the rule-base; and (3) context adaptation will not affect the sematic ordering of linguistic variables. Context adaptation of FISs has been carried using either transformation functions or adaptive operators. Most context adaptation studies focused on use of transformation functions (Gudwin and Gomide 1994; Gudwin et al. 1998; Magdalena 1997). According to Botta (2008), a transformation function serves to adapt a database (DB) or fuzzy partition, made up a group of fuzzy sets each representing a linguistic term, by mapping a normalized universe of discourse to the context adapted universe, possibly modifying the distribution (i.e., support and core), and the shape of fuzzy sets. Transformation functions have been commonly applied to adapt a base partition defined over a normalized universe of discourse $[0, 1]$, as shown in Fig. 6.2, where the universe of discourse is partitioned using common membership functions like triangular, trapezoidal or Gaussian (Ho 2013), to another context partition defined over a universe of discourse of $[a, b]$ as shown in Fig. 6.2.

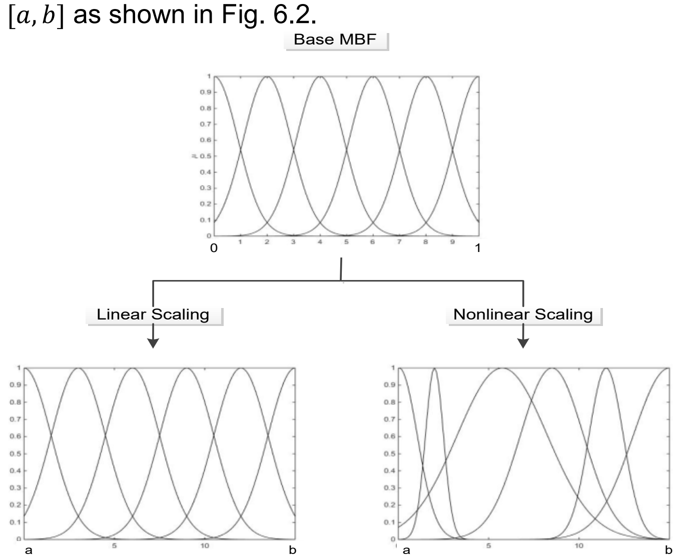

Figure 6.2: Fuzzy Inference System Context Adaptation Using Transformation Functions

The use of a normalized universe of discourse makes the database more general or context independent as the membership functions in the DB are defined over a normalized $[0, 1]$ range (Gudwin and Gomide 1994; Gudwin et al. 1998). Commonly, linear and nonlinear transformation functions are used. Linear transformation functions are applied on the overall partition of the fuzzy sets and will either linearly expand or contract all the fuzzy sets (see Fig. 6.2). However, nonlinear transformation functions are applied either to the overall partition or just on some points of the fuzzy sets (refer to Fig. 6.2), thus changing all fuzzy sets or the breakpoints (points of intersection) of selected fuzzy sets (Botta et al. 2009). The parameters of nonlinear transformation functions are derived using genetic algorithm or neural network based optimization approach over a data collected for the new or target context, using either field experiments or experts (Gudwin et al. 1998; Gudwin and Gomide 1994). Several nonlinear transformation functions have been applied. Magdalena (1997) proposed a sign function for nonlinear transformation of fuzzy sets and provided an application example using cart-pole balancing system control problem. Gudwin et al. (1998) used a linear combination of sigmoidal functions and demonstrated the development of the functions using assumed data. Pedrycz and Gomide (2007) suggested the use of piecewise linear function. According to Gudwin et al. (1998), transformation functions should meet certain requirements so as to preserve ordering and normality of the original linguistic variables. In particular, they are required to fulfil continuity, non-decreasing monotonicity, and boundary conditions. Furthermore, differentiability is expected when a learning algorithm is used to determine the optimal parameters of the transformation function.

Adaptive operators are specifically designed operators that adjust the universe of discourse of the fuzzy sets and modify the core, support, and shape of fuzzy sets. According to Botta (2008), the context adaptation process is based on a flexible nonlinear transformation function and four orthogonal fuzzy modifiers: core-position modifier, core-width modifier, support-width modifier, and membership function shape modifier, which are formulated using genetic algorithm based optimization process. The author tested the approach using four datasets including: a context-aware benchmarking dataset that arbitrarily assigns membership functions to a universe of discourse; structure of wage dataset including years of experience and wage; a synthetic dataset generated using a parametric function; and fuel consumption dataset. However, adaptive operators modify the partition in way that affects the order of the fuzzy sets and reduces the interpretability of the adapted partition, thus, limiting their direct application in model context adaptation (Botta et al. 2009).

The interpretability of the adapted fuzzy partitions can be verified using coverage, normality, and distinguishability properties (Botta et al. 2009). The adapted fuzzy partition has to cover the new context's universe of discourse $U = [a, b]$. The adapted membership functions of the fuzzy sets are also required to be normal so that at least one element of the universe of discourse will have full membership. The overlap of the fuzzy sets in the adapted fuzzy partitions has to be kept to a level that each couple of fuzzy sets are distinguishable enough (Botta 2008). The overlap between fuzzy sets A_1 and A_2 was measured using a possibility measure Π (Eq. 6.2). Possibility measure is recommended for efficient computation of distinguishability of fuzzy sets using overlap measure (Mencar et al. 2007). In order to ensure interpretability of the adapted fuzzy partition, the overlap among the fuzzy sets should not exceed 0.8 (Pulkkinen and Koivisto 2010).

$$\Pi(A_1, A_2) = \sup_{x \varepsilon U} min\{\mu_{A_1}(x), \mu_{A_2}(x)\} \tag{6.2}$$

Several context adaptation studies have been carried out (Gudwin et al. 1998; Magdalena 2002; Magdalena 1997; Botta 2008); however, certain limitations are observed. First, transformation function based studies were applied on normalized MFs defined over $[0, 1]$ range, resulting in the adaptation of theoretical context independent MFs to contexts-specific MFs. Second, most studies lacked practical application and rather focused on demonstrating the context adaptation method using benchmark datasets. Third, in fuzzy inference systems, not only the MFs but also the fuzzy operators and defuzzification methods are context-dependent (Klir and Yuan 1995). However, past studies mainly used minimum operator for input aggregation and implication, and centroid for defuzzification (Botta 2008), and did not evaluate the sensitivity of adapted models for fuzzy operators and defuzzification methods.

In this study, a context adaptation framework based on linear and nonlinear transformation functions have been adopted, for their ability to develop a transparent context adaptation framework for fuzzy inference based CLP models. The framework adapts the MFs and also tests the effect of fuzzy operators and defuzzification methods. The framework also focuses on adapting context-specific MFs that are defined over universe of discourse $B = [l, u]$ to another context defined over a universe of discourse $\mathrm{U} = [\mathrm{a}, \mathrm{b}]$. Additionally, the practical application of context adaptation is examined using construction labour productivity models developed for four unique construction contexts, where the context attributes that have been generated using the 5W1H (Who, What, Where, When, Why, and How) questions approach. The similarity of the adapted model with the original context-specific models is evaluated using agreement indices, and the usefulness of using the adapted models in predicting CLP is evaluated using the model's accuracies in terms of root mean square errors (RMSE).

6.3: CONTEXT ADAPTATION OF CLP MODELS USING TRANSFORMATION FUNCTIONS

The following procedure is developed for context-adaptation of fuzzy inference based CLP models. The underlying process in the context adaptation framework involves the determination of a context adaptation or transformation function using data from the new context for adapting the base A fuzzy set represented by the base MF to an adapted fuzzy set A' represented by the adapted MF (refer to Fig. 6.3).

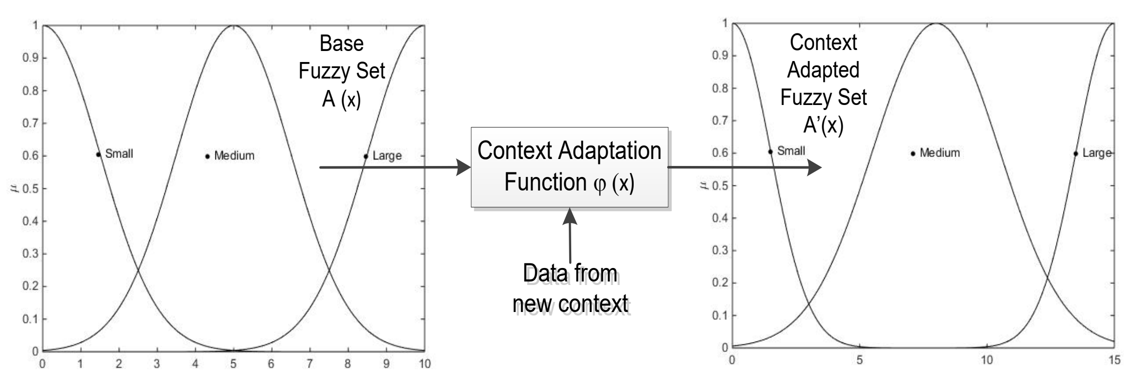

Figure 6.3: Context Adaptation Procedure

The procedure for context adaptation is summarized in the following steps:

1. Identify the base or original fuzzy sets for a given context and for each model feature (model input and output variables) of the base context-specific CLP model. The universe of discourse of each feature $\mathrm{B} = [\mathrm{l}, \mathrm{u}]$ and the parameters of the membership functions such as the standard deviation and modal values of the base fuzzy set $A(x, \sigma, \mu)$ are documented.
2. Collect data from the new context, using field data collection or experts, for the same features as the base model. The documented data set, $(\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_\mathrm{N})$, where N represents the total number of data instances of a given feature, which will then be used to determine the upper and lower limits of the adapted fuzzy set $\mathrm{A}'(\mathrm{x})$, the initial membership functions of the adapted fuzzy set $\mathrm{A_0}'(\mathrm{x})$, and the appropriate context adaptation function $\varphi(\mathrm{x})$.
3. Determine the boundary or upper and lower limits of the adapted fuzzy set $\mathrm{U} = [\mathrm{a}, \mathrm{b}]$ using the absolute limit context determination approach where the lower bound a is taken as the minimum data value or $a = \min(d_1, d_2, \ldots, d_N)$ and the upper bound b is taken as the maximum data value or $b = \max(d_1, d_2, \ldots, d_N)$.
4. Determine the initial membership functions of the adapted fuzzy set $\mathrm{A_0}'(\mathrm{x})$ using the collect data from the new context using either expert or data-driven membership function development approaches. The initial membership functions will be used to determine the nonlinear transformation function.
5. Determine the parameters of the context adaptation functions $\varphi_1(\mathrm{x})$ for linear and $\varphi_2(\mathrm{x})$ for nonlinear transformation of membership functions.
6. Develop the membership function of the adapted fuzzy set $\mathrm{A}'(\mathrm{x})$ for linear and nonlinear adaptation using Eq. (6.3):

$$A_i'(x) = A(\varphi_i(x)), x \in [a, b], \varphi_i(x) \in [l, u], i \in \{1,2\} \tag{6.3}$$

7. Adapt the base CLP model by replacing the base membership functions $A(x)$ with the adapted membership functions $\mathrm{A}'(\mathrm{x})$ for each model feature (input and output variables) and evaluate the prediction ability of the adapted CLP model. According to Ahmad et al. (2012), fuzzy inference model accuracies are commonly determined using the root mean square error (RMSE). The RMSE of the

adapted CLP model is calculated using Eq. (6.4), where $\mathrm{t_i}$ is the new context's target CLP value for the *i*th data instance, z_i is the corresponding predicted CLP value, and N is the total number of data instances.

$$RMSE_i = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(t_i - z_i)^2} \tag{6.4}$$

8. Evaluate the sensitivity of the adapted CLP model for fuzzy operators and defuzzification methods and summarize the improvement in prediction ability.
9. Determine the agreement between the adapted and base model of the new context using model agreement measures. The use of the modified Willmott agreement index is recommended to determine the similarity between models. The index is dimensionless, bounded, less sensitive to extreme values and outliers, and is suitable for cross-comparison between models (Willmott et al. 2012). The Willmott agreement index WI_i is shown in Eq. (6.5), where $\mathrm{P_i}$ represents the predicted values obtained from adapted model, $\mathrm{O_i}$ is the observed values obtained from the base model of the new context, $\overline{\mathrm{O}}$ is the mean value of the observed values, and N is the total number of data instances. The value WI_i of varies from 0 to 1 and a value of 1 indicates a perfect agreement between the adapted and base models.

$$WI_i = 1 - \frac{\sum_{i=1}^{N}|P_i - O_i|}{\sum_{i=1}^{N}(|P_i - \bar{O}| + |O_i - \bar{O}|)} \tag{6.5}$$

10. Compare and contrast the agreement indices and prediction ability of the resulting adapted models for either linear or nonlinear transformation of MFs, and combinations of fuzzy operators and defuzzification methods, and identify the most appropriate context adaptation approach.

6.4: Context Adaptation of Context-Specific Concreting Activity CLP Models

The preceding context adaptation procedure was tested using the field data collected for this research. In this study an operational definition of context for CLP modeling based on 5W1H (Who, What, Where, When, Why, and How) questions, which states that context is what constrains the four elements of a CLP model (user, model developer, model, and prevailing environment of the model) without intervening in the model development process explicitly, was adopted (refer to Section 5.2). Accordingly, four context-specific CLP models were developed and optimized for predicting labour productivity of concreting (concrete placement) activity using Mamdani-type fuzzy inference models; detailed discussion is provided in Section 5.6. The four context-specific models addressed concreting in Industrial, Warehouse, High-rise, and Institutional contexts. The properties of the CLP final models, summarized in Table 6.1, indicate the context-specific nature of models as they had distinctly different key influencing features (made up of factors, practices, and work sampling proportions), number of membership functions or number of rules, fuzzification coefficients, fuzzy operators, and defuzzification methods. The models were developed using Fuzzy C-Means clustering and Gaussian membership functions (MFs) were used.

Table 6.1: Base Context-Specific CLP Models: Features, Structure, and Model Parameters

Features, FIS structure, and model parameters	**Context**			
	1	**2**	**3**	**4**
	Concreting, Industrial buildings	**Concreting, Warehouse buildings**	**Concreting, High-rise buildings**	**Concreting, Institutional buildings**
Number of input features	16	7	8	11
Number of data instances	23	16	28	25
Model features	x3,x14,x45,x46,x49,x57, x106,x108,x129,x135, x138,x167,x168, y1,y2,y3,z	x45,x46,x48, x50,x108,y1, y7,z	x45,x46,x57,y1, y2, y3,y4,y6,z	x1,x5,x8,x18,x19, x46,x90,y1,y3,y4, y6,z
Fuzzification coefficient	1.5	2.5	2.0	2.0
Number of rules	6	7	6	7
Input aggregation operator	PROD	MIN	PROD	PROD
Implication method	PROD	MIN	PROD	PROD
Rule aggregation operator	MAX	SUM	PROBOR	PROBOR
Defuzzification method	MOM	BISECTOR	CENTROID	BISECTOR
Accuracy (RMSE)	1.162	0.467	0.992	0.671

The context adaptation framework presented in this chapter, adapts the four context-specific CLP models from one context to another, as shown in Fig. 6.4. Accordingly, the Industrial context CLP model will be adapted to suit the Warehouse, High-rise, and Institutional contexts using linear and nonlinear transformation functions, and a similar process will be repeated for the other contexts. The adaptation process for each context will result into six adapted models: three linearly adapted models and three nonlinearly adapted model (see Fig. 6.4). The adapted models will be compared with the base model developed for the given context (shown in Table 6.1) using modified Willmott agreement indices and model accuracy in terms of RMSE.

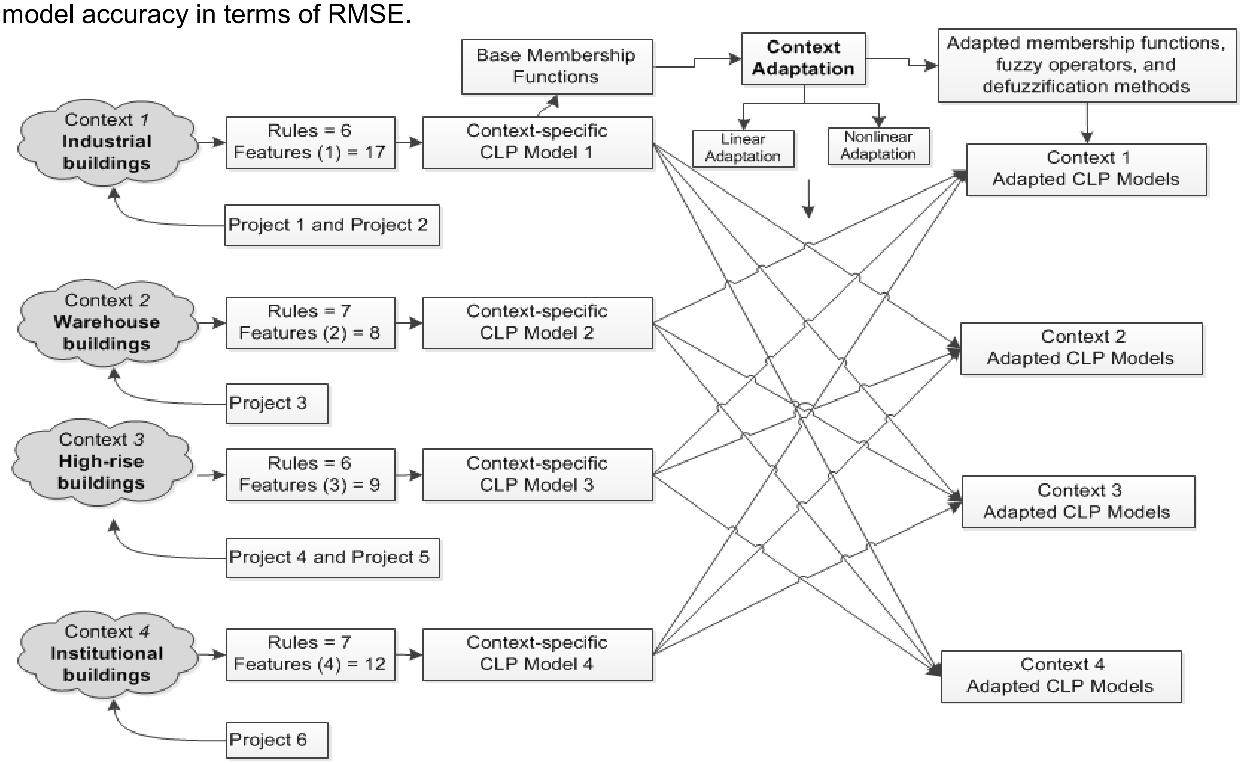

Figure 6.4: Context-Specific CLP Models Adaptation

As illustrated in the flowchart shown in Fig. 6.5, for given context-specific CLP model, all model membership functions are adapted using linear and nonlinear transformation functions to three other contexts, and initial model accuracies, $RMSE_{CA\text{-}L}$ and $RMSE_{CA\text{-}NL}$, respectively, are established. Then, the sensitivity of the adapted models for fuzzy operators and defuzzification methods are investigated for both linear and nonlinear adapted models and final model accuracies, $RMSE_{CA\text{-}LS}$ and $RMSE_{CA\text{-}NLS}$, respectively, are established. Next, the agreement indices (WI_i) between adapted and base models are computed for identifying the appropriate linear or nonlinear adaptation approach. Finally, the model accuracies of the adapted models are compared and the best performing context adaptation approach is identified.

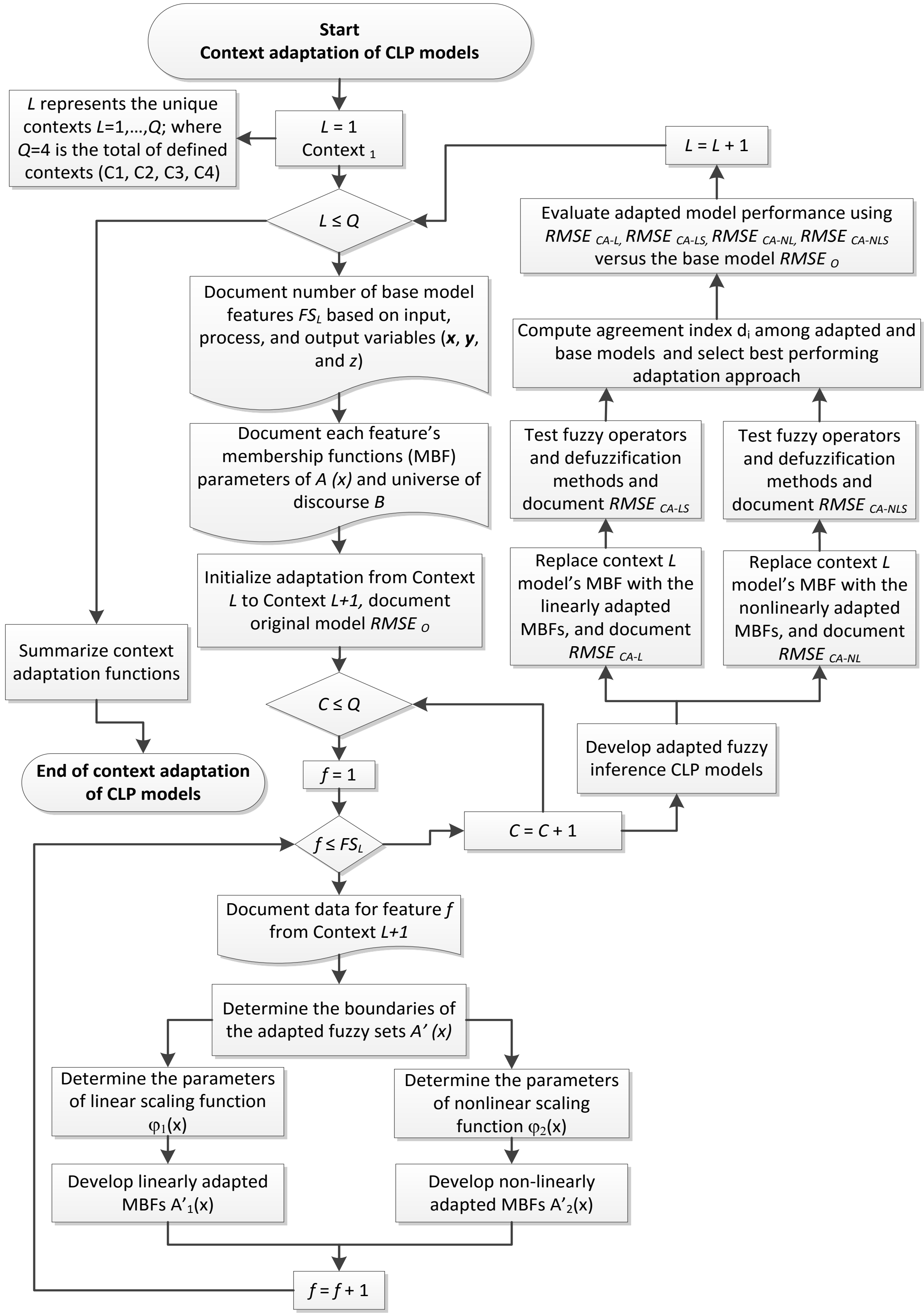

Figure 6.5: Flow Chart for Context Adaptation of CLP Models

In the following sections, a numerical illustration of the context adaptation framework is discussed based on the adaptation of the output or CLP feature from Context 4 (Institutional) to Context 3 (Industrial) context. The Institutional CLP model has 12 features and seven rules resulting in seven membership functions representing the fuzzy sets $A_j = \{A_1, A_2, \dots, A_7\}$. The base membership functions for CLP had a universe of discourse $B = [1.80, 11.25]$ and the mean CLP was 4.25 m^3/mhr, with standard deviation of 2.21. The membership functions were as shown in Fig. 6.6, where the following linguistic labels are used: $A_1(very-low)$, $A_2(low)$, $A_3(medium-low)$, $A_4(medium)$, $A_5(medium-high)$, $A_6(high)$, and $A_7(very-high)$. The parameters of the seven base CLP feature Gaussian membership functions $A(x,\sigma,\mu)$, where σ represents the spread of A and μ represents the modal value of A were: $A_1(x) = G(x, 0.973, 1.875)$, $A_2(x) = G(x, 0.573, 3.514)$, $A_3(x) = G(x, 0.786, 4.569)$, $A_4(x) = G(x, 0.864, 5.990)$, $A_5(x) = G(x, 0.877, 7.673)$, $A_6(x) = G(x, 0.707, 9.101)$, and $A_7(x) = G(x, 1.101, 11.105)$.

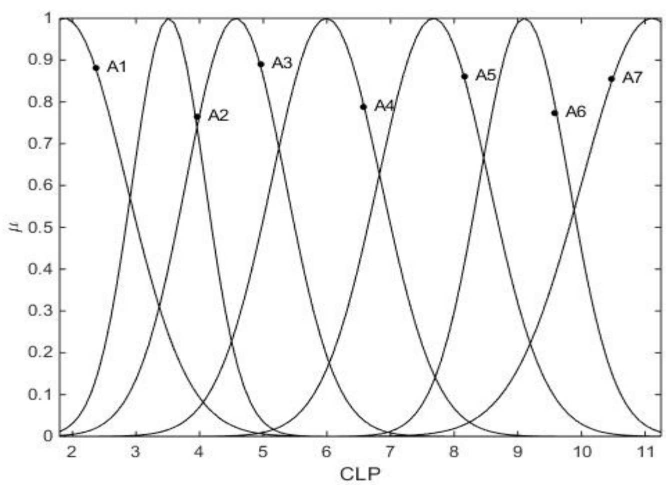

Figure 6.6: Base Membership Functions for CLP Feature in Institutional Context

The CLP data from Industrial context was then retrieved from the ProductivityTracker© database and had values of universe of discourse $U = [0.03, 8.66]$ and the mean CLP was 2.69 m^3/mhr, with standard deviation of 1.83. Next, the parameters of the context adaptation functions $\varphi_1(x)$ for linear adaptation and $\varphi_2(x)$ for nonlinear adaptation were developed as discussed in the following subsections.

6.4.1: Linear Adaptation

In linear adaptation, the base fuzzy sets defined over a universe of discourse of $B = [l, u]$ are adapted to the context-adapted universe of discourse by means of a linear transformation function shown in Eq. (6.5), where $U = [a, b]$ is used to represents the bounds of the adapted fuzzy sets:

$$\varphi_1(x, a, b) = \frac{(b - a)}{(u - l)} x + a \tag{6.5}$$

Accordingly, the respective Institutional CLP context linguistic variables $A(x)$ were adapted to Industrial CLP context linguistics variables $A'(x)$ using the linear transformation function show in Eq. (6.6):

$$\varphi_1(x) = x' = \frac{(8.66 - 0.03)}{(11.25 - 1.80)} x + 0.03 = 0.913x + 0.03 \tag{6.6}$$

Thus, the seven membership functions for CLP feature in the adapted context (Industrial context) $A'(x)$ are determined by replacing the parameters of the membership function $[\sigma, \mu]$ of the base fuzzy sets $A_j = \{A_1, A_2, \ldots, A_c\}$ with adapted values based on $\varphi_1(x)$. The parameters of the seven adapted CLP feature Gaussian membership functions $A'(x, \sigma', \mu')$ were: $A'_{11}(x) = G(x, 0.918, 1.742)$, $A'_{12}(x) = G(x, 0.553, 3.239)$, $A'_{13}(x) = G(x, 0.748, 4.203)$, $A'_{14}(x) = G(x, 0.819, 5.500)$, $A'_{15}(x) = G(x, 0.831, 7.123)$, $A'_{16}(x) = G(x, 0.676, 8.341)$, and $A'_{17}(x) = G(x, 1.035, 10.171)$.

The results of the linear adaptation of the fuzzy sets are shown in Fig. 6.7. The degree of overlap among the adapted sets was lower than the recommended maximum value of 0.8 (Pulkkinen and Koivisto 2010). Similar linear adaptation procedure was applied to the other 11 input variables or features of the Institutional CLP model. Then, the membership functions of the Institutional context model were replaced with the adapted ones, resulting in the linearly adapted Institutional context model for use in Industrial context. The linear adapted CLP model was used to predict CLP values of the Industrial context, which had 23 data instances as shown in Table 6.1. The adapted model had an initial $RMSE_{CA\text{-}L}$ value of 1.832.

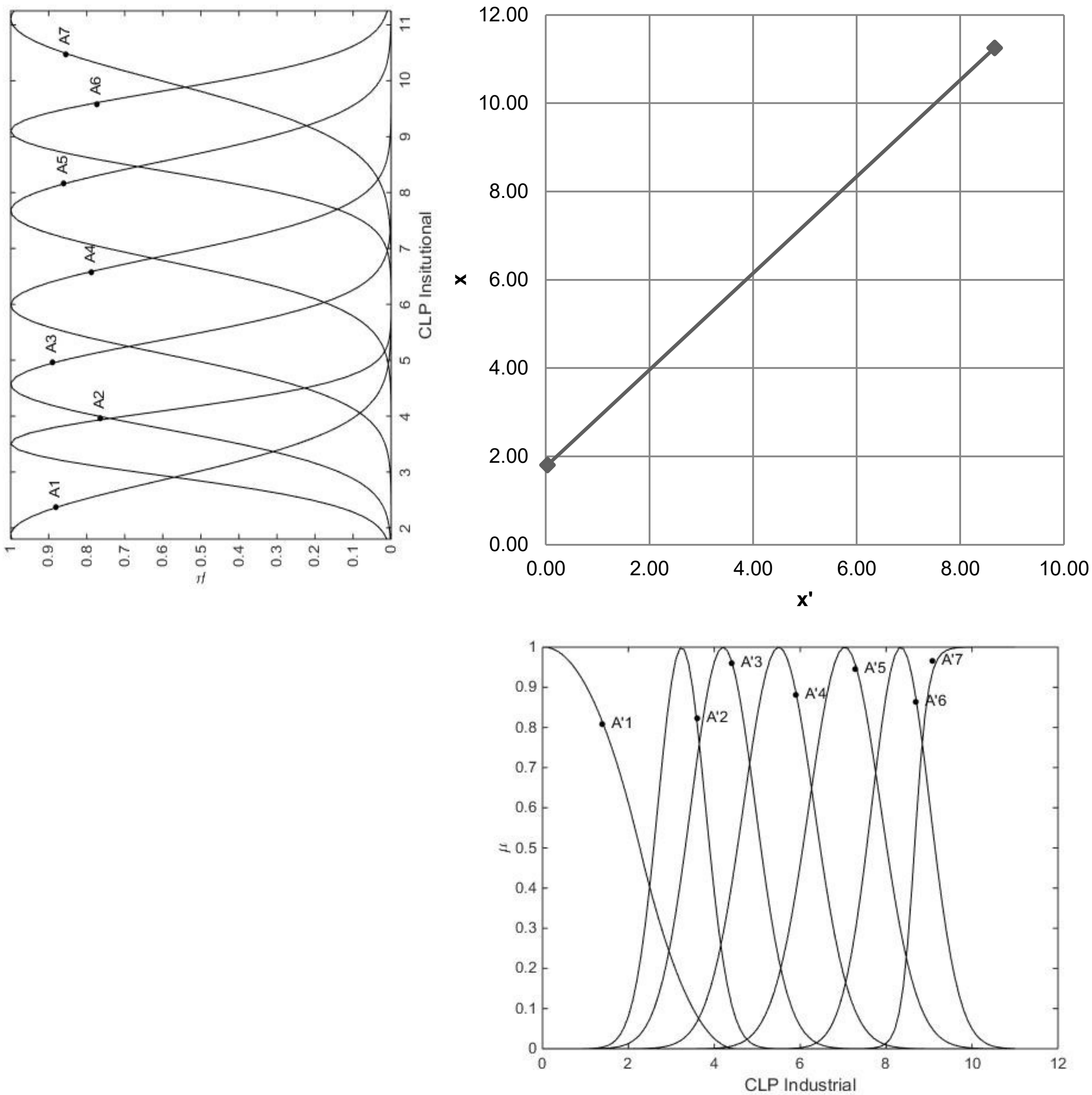

Figure 6.7: Linear Context Adaptation of CLP Feature from Institutional to Industrial Context

6.4.2: Nonlinear Adaptation

Nonlinear adaption involves the use of nonlinear transformation function that changes the universe of discourse of the base fuzzy sets, and also modifies the shape and distribution of the fuzzy sets in the space of the adapted universe of discourse (Botta 2008). In nonlinear adaptation, the base fuzzy sets defined over a universe of discourse of $B = [l, u]$ are adapted in the context-adapted universe of discourse by means of a nonlinear transformation function φ_2, where $U = [a, b]$ represents the identified bounds of the adapted fuzzy sets (refer to Fig. 6.8). The determination of the parameters of the nonlinear transformation function requires an optimization process (Gudwin et al. 1998).

In this research, a piecewise linear transformation function is used in order to develop an interpretable, logical, and fully invertible nonlinear context adaptation process (Pedrycz and Gomide 2007). For piecewise linear transformation function φ_2, shown in Fig. 6.8, the adjustable parameters $\boldsymbol{p}$ is made up of a collection of the split points $r_1, r_2, \dots, r_5$ and associated difference $D_1, D_2, \dots, D_5$; represented as $\mathbf{p} = [r_1, r_2, \dots, r_5, D_1, D_2, \dots, D_5]$. The piecewise functions will result in nonlinear mapping as some regions of x will be contracted and some of them will be expanded, resulting in modification of the shape and distribution of the fuzzy sets in the space of the adapted universe of discourse (Pedrycz and Gomide 2007).

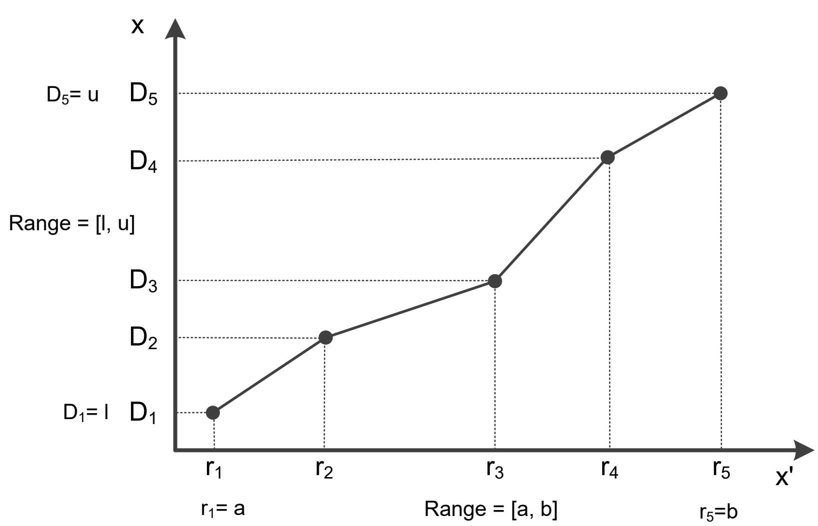

Figure 6.8: Nonlinear Context Adaptation Function: φ_2

In order to improve the optimization process required for determining the parameters of the piecewise linear transformation function, with limited data from new contexts, the number of split points has been kept to five points. According to Gudwin et al. (1998), context transformation functions are expected to fulfil the following requirements: continuity, monotonicity, and boundary conditions. The use of specifically non-decreasing monotonic functions ensured that the meaning and order of the linguistic terms is not changed (Pedrycz et al. 1997). Additionally, the boundary conditions $\varphi_2(\mathrm{l}) = \mathrm{a}$ and $\varphi_2(\mathrm{u}) = \mathrm{b}$ allowed the coverage of the new context data.

6.4.2.1: Computation of Nonlinear Transformation Function

Once the format or type of the nonlinear transfer function φ_2 is selected, the determination of the parameters of φ_2 in terms of $\mathbf{p} = [r_1, r_2, \ldots, r_5, D_1, D_2, \ldots, D_5]$ was carried out via optimization computations. The optimization process begins with the collection of fuzzy sets (linguistic terms) $A_j = \{A_1, A_2, \ldots, A_c\}$ as the base fuzzy sets of a given feature (model input or output variable) and the data set $(d_1, d_2, \ldots, d_N)$ collected for the same feature, but from the new context. Then, using Fuzzy C-Means (FCM) clustering, the initial membership functions of the adapted fuzzy sets $A_0{}' = \{A'_{10}, A'_{20}, \ldots, A'_{c0}\}$ were developed using the collected data set $(d_1, d_2, \ldots, d_N)$. The numbers of prototypes or cluster centres for FCM clustering are set equal to the number of base fuzzy sets, and the commonly used fuzzification coefficient of 2.0 is used. The resulting degrees of memberships of each data instances to the initial adapted fuzzy sets are arranged in the form of $N\ (c + 1)$ – tuples, as shown in Eq. (6.7), where the *k*th tuple consists of d_k that denotes some point in adapted universe of discourse x' where as $\mu_{k1}, \mu_{k2}, \ldots, \mu_{kc}$ are the numeric values of the corresponding membership degrees to initial adapted fuzzy sets $A'_{10}, A'_{20}, \ldots, A'_{c0}$, respectively:

$$\begin{array}{l} (d_1, (\mu_{11}, \mu_{12}, \ldots, \mu_{1c})) \\ (d_2, (\mu_{21}, \mu_{22}, \ldots, \mu_{2c})) \\ \ldots \\ (d_N, (\mu_{N1}, \mu_{N2}, \ldots, \mu_{Nc})) \end{array} \tag{6.7}$$

Then, the difference between the initial adapted degree of memberships and the degree of memberships computed using the transformation process was computed for each fuzzy set and for the respective data instance. This difference between the initial adapted fuzzy sets and the adapted fuzzy sets developed using the nonlinear transformation functions formed the objective function of the optimization process. The objective function Q was calculated using sum of squared errors as shown in Eq. (6.8):

$$Q(p) = \sum_{i=1}^{c}(A_i(\varphi(d_1, \boldsymbol{p}) - \mu_{1i})^2 + \sum_{i=1}^{c}(A_i(\varphi(d_2, \boldsymbol{p}) - \mu_{2i})^2 + \cdots + \sum_{i=1}^{c}(A_i(\varphi(d_N, \boldsymbol{p}) - \mu_{Ni})^2 \qquad (6.8)$$

Thus, the determination of the nonlinear transformation function involved the minimization of the objective function with respect to parameters of $\boldsymbol{p}$. The solution of this constrained nonlinear minimization optimization problem can be effectively developed using genetic algorithm, as the objective function is nonlinear and nonconvex (has multiple feasible regions), and use of gradient based optimization techniques, such as generalized reduced gradient approach, will only lead to local optimum solutions (Gudwin et al. 1998). Accordingly, an optimization process using genetic algorithm was used, and the parameters of the nonlinear transformation function $\varphi_2 = \mathbf{p} = [r_1, r_2, \ldots, r_5, D_1, D_2, \ldots, D_5]$ were used for real coding of the chromosome in the genetic optimization process. The objective of the genetic search was to minimize the objective function Q, and the fitness value of each solution was determined by calculating Q. Then, the genetic operations of reproduction, crossover, and mutation were performed. Each operation generated new sets of chromosomes, representing a new solution that meets the optimization constraints. The solution chromosomes are checked according to the following nonlinear context adaptation constraints: (1) The parameters of the nonlinear transformation function must be greater than zero; (2) Boundary conditions for coverage of the new context data, defined over $U = [a, b]$, and using the base fuzzy set range of $B = [l, u]$: $r_1 = \mathrm{l}$, $r_5 = u$, $D_1 = a$, and $D_5 = b$; (3) The *j^th^* split value r_j must not be greater than that of the *j+1^th^* split value r_{j+1}, and (4) The *j^th^* difference value D_j must not be greater than that of the *j+1^th^* difference value D_{j+1}. The genetic optimization was implemented using Microsoft Excel Solver (Frontline 2010). Initial population of 100 solutions was randomly generated. A mutation rate of 0.075 and stopping criteria based on convergence value of 0.0001 was used.

6.4.2.1: Computational Results for Nonlinear Adaptation of CLP Models

The numerical illustration of the nonlinear context adaptation of the output or CLP feature from Context 4 (Institutional) to Context 3 (Industrial) building context is discussed here. As discussed above, the seven base fuzzy sets $A_j = \{A_1, A_2, \ldots, A_7\}$ for CLP feature and associated membership functions from Institutional context were first documented. Next the CLP data set $(d_1, d_2, \ldots, d_N)$ collected from the Industrial context is retrieved from the ProductivityTracker© database and had values of universe of discourse of $U = [0.03, 8.66]$. Using Fuzzy C-Means (FCM) clustering, the initial membership functions of the adapted fuzzy sets $A'_0 = \{A'_{01}, A'_{02}, \ldots, A'_{07}\}$ were developed, where the number of prototypes for FCM clustering was set at seven. Then, using genetic algorithm based optimization, the parameters of the nonlinear piecewise transformation function $\mathbf{p}$ were developed, and the resulting parameters were $\varphi_2 = [r_1 = 0.03, r_2 = 2.05, r_3 = 5.91, r_4 = 6.36, r_5 = 8.66, D_1 = 1.80, D_2 = 5.52, D_3 = 7.76,\ D_4 = 9.63, D_5 = 11.25]$.

Thus, the seven membership functions for CLP feature in the adapted context (Industrial context) $A'_{2j}(x) = \{A'_{21}, A'_{22}, \ldots, A'_{2c}\}$ were determined by replacing the parameters of the membership function $[\sigma, \mu]$ of the base fuzzy sets $A_j = \{A_1, A_2, \ldots, A_c\}$ with adapted values based on $\varphi_2(x)$. The results of the nonlinear adaptation of the fuzzy sets are shown in Fig. 6.9. The parameters of the seven nonlinearly adapted CLP feature Gaussian membership functions $A'_{2j}(x, \sigma', \mu')$ were: $A'_{21}(x) = G(x', 0.528, 1.018)$, $A'_{22}(x) = G(x, 0.311, 1.908)$, $A'_{23}(x) = G(x, 0.427, 2.481)$, $A'_{24}(x) = G(x,\ 0.469, 2.860)$, $A'_{25}(x) = G(x, 0.476, 5.760)$, $A'_{26}(x) = G(x, 0.384, 6.233)$, and $A'_{27}(x) = G(x, 0.598, 8.454)$.

The results shown in Fig. 6.9 indicate that the shape and distribution of the fuzzy sets have been modified, and in some case the degree of overlap among the adapted sets was higher than the recommended maximum value of 0.8 (Pulkkinen and Koivisto 2010). This will naturally reduce the interpretability of the adapted models, however, this is a common problem witnessed in nonlinear adaptation of fuzzy systems (Ho 2013).

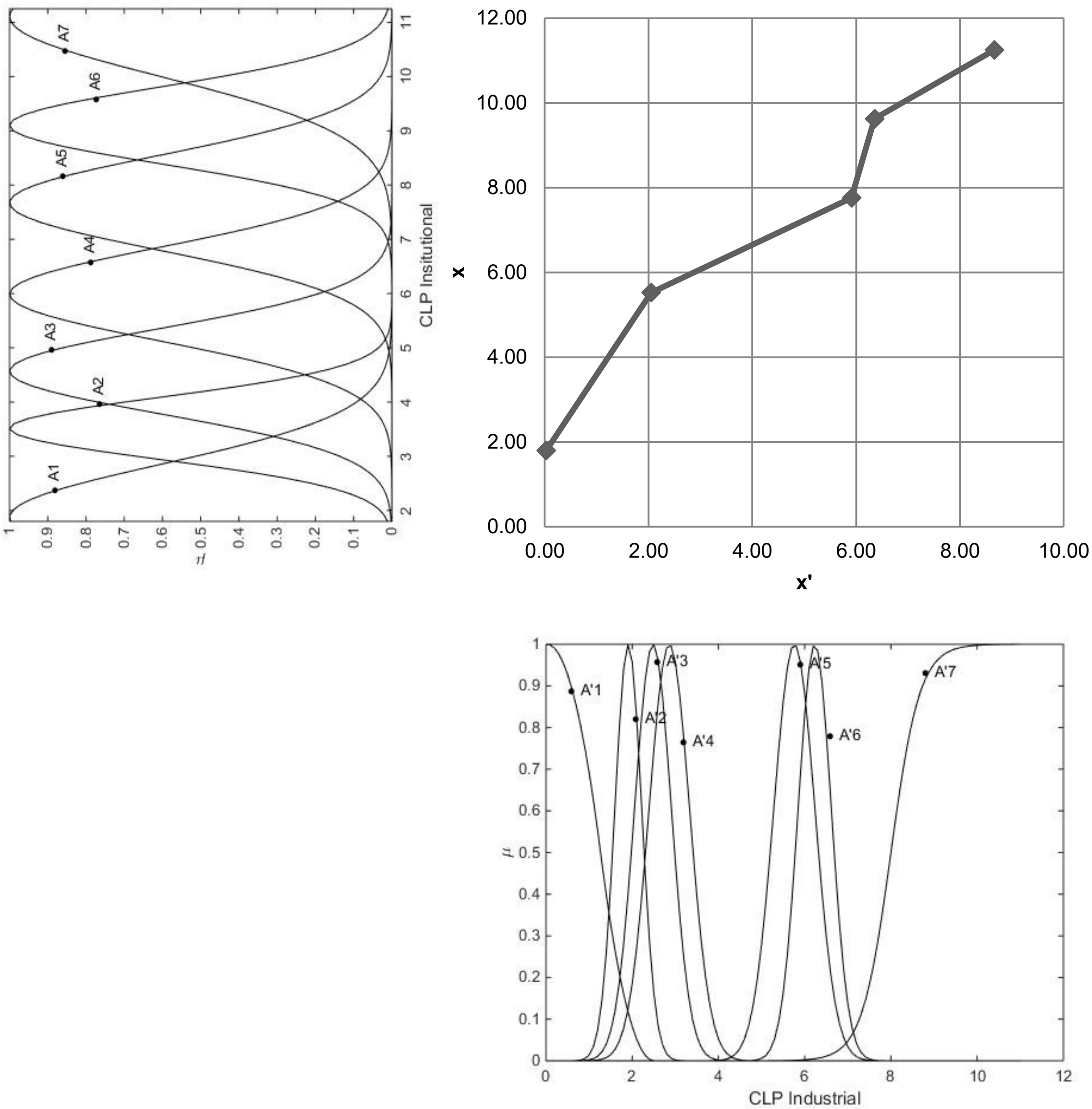

Figure 6.9: Nonlinear Context Adaptation of CLP Feature from Institutional to Industrial Context

Similar nonlinear adaptation procedure was applied to the remaining 11 input variables or features of the Institutional CLP model. Then, the membership functions of the Institutional context model were replaced with the adapted ones, resulting in the nonlinearly adapted Institutional context model for use in Industrial context. The nonlinearly adapted CLP model was used to predict CLP values of the

Industrial context, which had 23 data instances as shown in Table 6.1. The adapted model had an initial $RMSE_{CA\text{-}NL}$ value of 2.742.

6.4.3: Sensitivity Analysis of Adapted Models for Fuzzy Operators and Defuzzification Methods

The sensitivity of the linearly and nonlinearly adapted CLP models was then further evaluated by changing the fuzzy operators and defuzzification methods. The following options of fuzzy operators and defuzzification methods were tested: for input aggregation [MIN (minimum) and PROD (product)], for implication [MIN (minimum) and PROD (product)], for rule aggregation [MAX (maximum), SUM (sum of each rule's output set), and PROBOR (probabilistic OR)], and for defuzzification [CENTROID, BISECTOR, MOM (middle of maximum), LOM (largest of maximum), and SOM (smallest of maximum)]. The sensitivity options were varied one at a time, and a total of 30 unique combinations were tested. The options and results for Context 4 model, which was linearly adapted to suit Context 1, are shown in Table 6.2.

For each linear and nonlinear adapted CLP model, the final adapted model accuracy measures $RMSE_{CA\text{-}LS}$ and $RMSE_{CA\text{-}NLS}$ of 1.832 and 2.738, respectively, were determined based on the best combination of the listed options of fuzzy operators and defuzzification methods. The results yielding the lowest $RMSE_{CA\text{-}LS}$ and $RMSE_{CA\text{-}NLS}$ values provided the best adapted fuzzy operators and defuzzification methods for CLP models. Finally, the agreement index WI_i between the adapted and the base model was computed using the modified Willmott index (Eq. 6.4). Accordingly, the appropriate context adaptation approach was the linear adapted model, which had an agreement index of 0.340, as the nonlinear adapted model had an agreement index of only 0.142.

6.4.4: Results and Discussion

The linear and nonlinear context adaptation framework was applied on the four context-specific CLP models as shown in Fig. 6.4. The results of the context adaptation process are summarized for each context, and are discussed in the following sections.

Table 6.2: Context Adapted Models Sensitivity Analysis

Case	Fuzzy operators and Defuzzification methods				
	Input aggregation	**Implication method**	**Rule aggregation**	**Defuzzification method**	**Accuracy (RMSE)**
1	MIN	MIN	MAX	CENTROID	1.840
2	MIN	MIN	MAX	BISECTOR	1.840
3	MIN	MIN	MAX	MOM	1.955
4	MIN	MIN	MAX	LOM	2.030
5	MIN	MIN	MAX	SOM	1.955
6	MIN	MIN	SUM	CENTROID	1.840
7	MIN	MIN	SUM	BISECTOR	1.840
8	MIN	MIN	SUM	MOM	1.955
9	MIN	MIN	SUM	LOM	2.030
10	MIN	MIN	SUM	SOM	1.955
11	MIN	MIN	PROBOR	CENTROID	1.840
12	MIN	MIN	PROBOR	BISECTOR	1.840
13	MIN	MIN	PROBOR	MOM	1.955
14	MIN	MIN	PROBOR	LOM	2.030
15	MIN	MIN	PROBOR	SOM	1.955
16	PROD	PROD	MAX	CENTROID	**1.832**
17	PROD	PROD	MAX	BISECTOR	1.840
18	PROD	PROD	MAX	MOM	1.955
19	PROD	PROD	MAX	LOM	2.030
20	PROD	PROD	MAX	SOM	1.955
21	PROD	PROD	SUM	CENTROID	**1.832**
22	PROD	PROD	SUM	BISECTOR	1.840
23	PROD	PROD	SUM	MOM	1.840
24	PROD	PROD	SUM	LOM	2.030
25	PROD	PROD	SUM	SOM	1.955
26	PROD	PROD	PROBOR	CENTROID	**1.832**
27	PROD	PROD	PROBOR	BISECTOR	1.840
28	PROD	PROD	PROBOR	MOM	1.955
29	PROD	PROD	PROBOR	LOM	2.030
30	PROD	PROD	PROBOR	SOM	1.955

6.4.4.1: Industrial Context CLP Models

For the Industrial context, six adapted models were developed from Warehouse, High-rise, and Institutional contexts and based on linear and nonlinear adaptation. The base CLP models for Industrial context had an RMSE value of 1.162. In Table 6.3, the initial RMSE values represent the accuracy of adapted models in predicting the CLP values of the Industrial context and the final RMSE values represent the accuracy of the adapted models after sensitivity analysis. The best performing fuzzy operators and defuzzification methods are also shown in Table 6.3. Based on the agreement indices of the six adapted models, as shown in Table 6.3, the model linearly adapted from Institutional context has the highest agreement to the base model of the Industrial context, with an agreement index value of 0.340. The linearly adapted model from the Institutional context has an RMSE value of 1.832. Additionally, the linear adapted models from all three contexts, as compared to the nonlinear adapted models, had better agreement with the base model and also had higher accuracy based on RMSE values. However, in terms of model accuracy, none of the adapted models performed better than the base model.

Table 6.3: Context Adaptation Results for Industrial Context (C1)

Adapted Models	**Adapted from Warehouse:C2**		**Adapted from High-rise:C3**		**Adapted from Institutional:C4**	
	Linear	**Nonlinear**	**Linear**	**Nonlinear**	**Linear**	**Nonlinear**
RMSE (initial)	1.892	2.079	1.880	1.880	1.832	2.742
RMSE (final)	1.831	2.028	1.873	1.873	1.832	2.738
Sensitivity improvement (%)	3.26	2.47	0.32	0.32	0.00	0.14
Agreement index	0.203	0.188	0.282	0.282	**0.340**	0.142
Input aggregation operator	PROD	PROD	PROD	PROD	PROD	PROD
Implication method	PROD	PROD	PROD	PROD	PROD	PROD
Rule aggregation operator	SUM	MAX	PROBOR	PROBOR	PROBOR	PROBOR
Defuzzification method	BISECTOR	MOM	MOM	MOM	CENTROID	CENTROID

6.4.4.2: Warehouse Context CLP Models

For the Warehouse context, six adapted models were developed from Industrial, High-rise, and Institutional contexts and based on linear and nonlinear adaptation. The base CLP models for Warehouse context had an RMSE value of 0.467. In Table 6.4, the initial RMSE values represent the accuracy of adapted models in predicting the CLP values of the Warehouse context and the final RMSE values represent the accuracy of the adapted models after sensitivity analysis. The best performing fuzzy operators and defuzzification methods are shown in Table 6.4. Based on the agreement indices of the six adapted models, as shown in Table 6.4, the model nonlinearly adapted from Industrial context has the highest agreement to the base model of the Warehouse context, with an agreement index value of 0.459. The nonlinearly adapted model from the Industrial context has an RMSE value of 0.939. In terms of model accuracy, none of the adapted models performed better than the base model.

Table 6.4: Context Adaptation Results for Warehouse Context (C2)

Adapted Models	**Adapted from Industrial:C1**		**Adapted from High-rise:C3**		**Adapted from Institutional:C4**	
	Linear	**Nonlinear**	**Linear**	**Nonlinear**	**Linear**	**Nonlinear**
RMSE (initial)	1.637	0.946	0.731	0.753	2.282	3.286
RMSE (final)	1.610	0.939	0.719	0.753	2.267	3.264
Sensitivity improvement (%)	1.68	0.74	1.67	0.00	0.69	0.68
Agreement index	0.201	**0.459**	0.214	0.349	0.144	0.115
Input aggregation operator	PROD	PROD	PROD	PROD	PROD	PROD
Implication method	PROD	PROD	PROD	PROD	PROD	PROD
Rule aggregation operator	MAX	MAX	MAX	MAX	PROBOR	PROBOR
Defuzzification method	CENTROID	CENTROID	MOM	CENTROID	MOM	MOM

6.4.4.3: High-rise Context CLP Models

For the High-rise context, six adapted models were developed from Industrial, Warehouse, and Institutional contexts and based on linear and nonlinear adaptation. The base CLP models for High-rise context had an RMSE value of 0.992. In Table 6.5, the initial RMSE values represent the accuracy of adapted models in predicting the CLP values of the High-rise context and the final RMSE values represent the accuracy of the adapted models after sensitivity analysis. The best performing fuzzy operators and defuzzification methods are shown in Table 6.5. Based on the agreement indices of the six adapted models, as shown in Table 6.5, the model nonlinearly adapted from Warehouse context has the highest agreement to the base model of the Industrial context, with an agreement index value of 0.427. The nonlinearly adapted model from the Warehouse context has an RMSE value of 3.851. Additionally, the nonlinear adapted models from all three contexts, as compared to the linear adapted models, had better or equal agreement with the base model. However, both linearly and nonlinear adapted models have similar accuracy based on RMSE values, and none of the adapted models performed better than the base model.

Table 6.5: Context Adaptation Results for High-rise Context (C3)

Adapted Models	Adapted from Industrial:C1		Adapted from Warehouse:C2		Adapted from Institutional:C4	
	Linear	**Nonlinear**	**Linear**	**Nonlinear**	**Linear**	**Nonlinear**
RMSE (initial)	3.797	3.797	4.350	4.128	4.627	4.627
RMSE (final)	3.379	3.379	3.851	3.851	4.627	4.627
Sensitivity improvement (%)	11.01	11.01	11.46	6.71	0.00	0.00
Agreement index	0.048	0.048	0.395	**0.427**	0.181	0.181
Input aggregation operator	PROD	PROD	MIN	MIN	PROD	PROD
Implication method	PROD	PROD	MIN	MIN	PROD	PROD
Rule aggregation operator	MAX	MAX	SUM	SUM	PROBOR	PROBOR
Defuzzification method	MOM	MOM	BISECTOR	BISECTOR	BISECTOR	BISECTOR

6.4.4.4: Institutional Context CLP Models

For the Institutional context, six adapted models were developed from Industrial, Warehouse, and High-rise contexts and based on linear and nonlinear adaptation. The base CLP models for Institutional context had an RMSE value of 0.671. In Table 6.6, the initial RMSE values represent the accuracy of adapted models in predicting the CLP values of the Institutional context and the final RMSE values represent the accuracy of the adapted models after sensitivity analysis. The best performing fuzzy operators and defuzzification methods are shown in Table 6.6. Based on the agreement indices of the six adapted models, as shown in Table 6.6, the model linearly adapted from Industrial context has the highest agreement to the base model of the Institutional context, with an agreement index value of 0.398. The linearly adapted model from the Institutional context has an RMSE value of 2.552. In terms of model accuracy, none of the adapted models performed better than the base model.

Table 6.6: Context Adaptation Results for Institutional Context (C4)

Adapted Models	**Adapted from Industrial:C1**		**Adapted from Warehouse:C2**		**Adapted from High-rise:C3**	
	Linear	**Nonlinear**	**Linear**	**Nonlinear**	**Linear**	**Nonlinear**
RMSE (initial)	2.552	2.212	3.343	3.083	2.950	3.609
RMSE (final)	2.552	2.213	2.530	2.634	2.945	3.446
Sensitivity improvement (%)	0.00	0.00	24.33	14.57	0.17	4.52
Agreement index	**0.398**	0.199	0.329	0.357	0.383	0.362
Input aggregation operator	PROD	PROD	MIN	PROD	MIN	MIN
Implication method	PROD	PROD	MIN	PROD	MIN	MIN
Rule aggregation operator	MAX	MAX	SUM	MAX	MAX	MAX
Defuzzification method	CENTROID	MOM	BISECTOR	MOM	BISECTOR	MOM

In summary, the review of the linear and nonlinear adaptation approaches indicated that, linear adapted models were in agreement with that of the base models for Industrial and Institutional contexts, and nonlinear adapted models were in agreement with that of the base models for Warehouse and High-rise contexts. However, the use of nonlinear adaptation approach has resulted in reduced interpretability of the adapted models. In terms of prediction accuracy, in all four contexts, none of the adapted models performed better than the base models. This is expected as the base models have been developed and further optimized using the key context-specific variables influencing CLP and the associated dataset. Additionally, the sensitivity analysis on fuzzy operators and defuzzification methods did not show significant improvement in adapted model's accuracy. The comparison of the adapted models accuracy with the accuracy of the original context-specific models before optimization showed promising results for the Industrial and Warehouse contexts. The original Industrial context model before optimization had an

RMSE value of 1.582 and the most accurate linearly adapted model from Warehouse context has an RMSE value of 1.831. Similarly, the original Warehouse context model before optimization had an RMSE value of 0.586 and the most accurate linearly adapted model from High-rise context has an RMSE value of 0.719. Considering the effort required for collecting data on all influencing variables and to develop and optimize new models, the use of context adaptation framework presented in this chapter, which enables the reuse of existing CLP models, has provided a simpler and efficient alternative for developing CLP models for the Industrial and Warehouse contexts. Additionally, context adaptation enables the transfer of knowledge from one context to another. Accordingly, based on the results discussed above, the following contexts are found to be similar to one another. For Industrial context, Institutional context was found to be the closest one for model adaptation purposes, which is followed by the High-rise context. For Warehouse context, Industrial context was found to be the closest one for model adaptation purposes, which is followed by the High-rise context. For High-rise context, Warehouse context was found to be the closest one for model adaptation purposes, which is followed by the Institutional context. For Institutional context, Industrial context was found to be the closest one for model adaptation purposes, which is followed by the High-rise context.

6.4: Chapter Summary

Construction labour productivity is one of the most studied areas in the construction research field, and several predictive models have been developed; however, a framework for adapting the several models from one context to another is missing. This chapter presents a context adaptation framework for transferring the knowledge represented in CLP models from one context to another. The chapter has provided the background and review of existing approaches for context adaptation of fuzzy systems. Limitations of past studies were also summarized. A general procedure for linear and nonlinear adaptation of context-specific CLP models was formulated based on the transformation of the membership functions, and further sensitivity analysis of adapted models using fuzzy operators and defuzzification methods.

Subsequently, using four context-specific CLP models for concreting activity under Industrial, Warehouse, High-rise, and Institutional building contexts, the developed context adaptation framework was implemented. The results indicated that linear adaptation approach was in agreement with the base model in two contexts and the nonlinear adaptation approach was in agreement with the base model in the other two contexts. However, in terms of model accuracy, none of the adapted models performed better than the base models of a given context. Additionally, the sensitivity analysis on fuzzy operators and defuzzification methods did not show significant improvement in adapted model accuracy. Furthermore, the best adapted model for each context was reviewed and contextual similarities in terms of CLP prediction were examined. In the next chapter, the framework for the development of universal CLP models is presented. The framework describes the processes involved in the abstraction of the context-specific models, so as to develop a generalized CLP model.

6.5: References

Ahmad, S. S. S., Ahmad, S., and Pedrycz, W. (2012). “Data and feature reduction in fuzzy modeling through particle swarm optimization.” *Appl. Comput. Intell. Soft Comput.,* 2012, 21 pp.

Awad, A., and Fayek, A. R. (2013). "Adaptive learning of contractor default prediction model for surety bonding." *J. Constr. Eng. Manage*., 139(6), 694–704.

Botta, A. (2008). “Automatic context adaptation of fuzzy systems.” Ph.D. thesis, IMT Institute for Advanced Studies, Lucca, Italy.

Botta, A., Lazzerini, B., Marcelloni, F., and Stefanescu, D. C. (2009). "Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index." *Soft Compu.*, 13(5), 437–449.

Cordón, O., Herrera, F., Magdalena, L., and Villar, P. (2001). "A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base." *Inf.Sci.*, 136(1), 85–107.

Engwall, M. (2003). "No project is an island: linking projects to history and context." *Res. Pol.,* 32(5), 789–808.

Frontline (2015). “Introduction: Genetic algorithms and evolutionary algorithms.” <http://www.solver.com/genetic-algorithm> (March 16, 2015).

Gudwin, R.R., and Gomide, F. (1994). "Context adaptation in fuzzy processing." *Proc., Brazil-Japan Joint Symposium on Fuzzy Systems*, Campinas, Brazil.

Gudwin, R. R., Gomide, F. A., and Pedrycz, W. (1998). "Context adaptation in fuzzy processing and genetic algorithm." *Int. J. Intell. Syst*., 13(10-11), 929–948.

Ho, D. T. (2013). “Context dependent fuzzy modelling and its application.” PhD thesis, Univ. of Nottingham, Nottingham, UK.

Ji, S., Park, M., and Lee, H. (2012). "Case adaptation method of case-based reasoning for construction cost estimation in Korea." *J. Constr. Eng. Manage*., 138(1), 43–52.

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

Magdalena, L. (2002). "On the role of context in hierarchical fuzzy controllers." *Int. J. Intell. Syst.,* 17(5), 471–493.

Magdalena, L. (1997). "Adapting the gain of an FLC with genetic algorithms." *Int. J. Approx. Reasoning,* 17(4), 327–349.

Oduba, A. O. (2002). “Predicting industrial construction productivity using fuzzy expert systems.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Pedrycz, W., and Gomide, F. (2007). *Fuzzy systems engineering: Toward human-centric computing*. John Wiley and Sons, Hoboken, New Jersey, US.

Pedrycz, W., Gudwin, R., and Gomide, F. (1997). "Nonlinear context adaptation in the calibration of fuzzy sets." *Fuzzy Sets Syst*., 88(1), 91–97.

Pulkkinen, P., and Koivisto, H. (2010). "A dynamically constrained multiobjective genetic fuzzy system for regression problems." *IEEE T. Fuzzy Syst.*, 18, 161–177.

Willmott, C., Robeson, S., and Matsuura K. (2012). “A refined index of model performance.” *Int. J. Climatol.*, 32, 2088–2094.

Yi, W., and Chan, A. (2014). "Critical review of labor productivity research in construction journals." *J. Manage. Eng.,* 30(2), 214–225.

Chapter 7: Developing and Optimizing Universal CLP Models[7]

7.1: INTRODUCTION AND BACKGROUND

Construction labour productivity (CLP) deals with the efficiency of crews in the complex conversion process of inputs (labour, material, equipment, etc.) to outputs (project products) in construction projects. Numerous objective and subjective factors, practices, and work sampling proportions are known to affect the process. CLP, computed as a ratio of output to input, accounts for one-third to one-half of overall project costs (Hanna et al. 2005). However CLP still remains the major source of project risk and exhibits the highest variability among project resources, and research is being carried out to understand, model, and formulate improvement strategies (Tsehayae and Fayek 2014).

Modeling CLP remains a challenge as the influencing variables (factors, practices, and work sampling proportions) are numerous, complex, dynamic, and inconsistent from project to project. Several models have been developed based on different data analysis methods to determine the relationships between mainly objective input variables like temperature and output in terms of labour productivity (Yi and Chan 2014). The most common models are based on multi-linear regression analysis. For example, models for masonry wall activity were developed using data from three projects (Thomas et al. 1990). A study on housing project showed that work sampling proportions could be used to indicate actual site productivity and crew learning rates (Handa and Abdalla 1989). Since the end of 1990s, CLP studies have focused on the use of artificial neural networks (NN) and have been used for concrete formwork (Portas and AbouRizk 1997), and formwork assembly, steel fixing, and concrete pouring/finishing operations (Ezeldin and Sharara 2006). Expert systems were also used for masonry construction (Hendrickson et al. 1987) and fuzzy inference systems were used for industrial pipe rigging and welding activities (Fayek and Tsehayae 2012). However, the models developed for the previously encountered projects suit only a specific context based on project type, nature of activity under study, and external environment (weather, contract type, project location, etc.) and their implementation is mostly restricted to

---

[7] Parts of this chapter have been published in the Proceedings, IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS) 2013, Edmonton, Alberta, Canada, June 24-28, pp. 1096-1101.

the data used in their development. In addition, most models were also not able to deal with subjective variables in a comprehensive manner.

A general, more abstract and universal CLP model, which is not heavily dependent on the often context-specific existing models, but is based on the important knowledge captured in the respective models is non-existent in construction research (Tsehayae et al. 2013). In one of the few labour productivity studies that investigated more than one context-specific models, Mosehli et al. (2005) developed a software platform having five change order impact evaluation models: two regression based models for general construction (Moselhi et al. 1991 and Ibbs 1997); a regression based models for mechanical activities (Hanna et al. 1999a); a regression based models for electrical activities (Hanna et al. 1999b); and a neural network model for general construction (Mosehli et al. 2005). However, the models were used for comparison purposes only and the suitability and choice for use of the included models was left for the user to decide. In a study to identify the levels of wasted time proportion in construction projects, a meta-analysis based methodology was used to provide a synthesis of the findings of 22 productivity studies, and the analysis revealed that an average of 49.6% of time is wasted in construction projects (Horman and Kenley 2005). According to the authors, meta-analysis was used to collect relevant statistics from individual studies and combine them to come up with an average result. Meta-analysis provides useful findings; however, the approach is not suited for prediction purposes, and the results are confined to statistical values such as averages, standard deviations, and range of values.

A universal CLP model will make use of information granules (i.e., groups, classes, intervals, or clusters of the data) in the process of modeling CLP (Yao 2007). An information granule can be defined as “points drawn together by indistinguishability, similarity, proximity or functionality” (Zadeh 1997) and fuzzy sets through the use of linguistic terms offer an important and unique feature for describing information granules. Information granularity and granular modeling are key in modeling complex, ill-defined systems where a mathematical model may be difficult or impossible to build (Dick and Kandel 2001); additionally, they have the capability to abstract context-specific models, which could come from different perspectives or contexts (Pedrycz and Song 2012). Thus, by extracting the most important relationships contained in the context-specific models, while suppressing fine details, universal models provide a realistic estimate of CLP and a better representation of the complex dynamic conversion process (Pedrycz et al. 2012).

The objective of this chapter is to develop a universal CLP model based on the abstraction of the four context-specific CLP knowledge bases for concreting (concrete placement) activity. The universal CLP model will represent a generalized context-free knowledge base that can be used to predict labour productivity of concreting activities in any context, as developing an adequate number of context-specific CLP models representing each unique construction context is difficult to achieve. First, the details of the four context-specific, also referred to as lower-level models, are presented. Then, the development of the universal model input and output data using the input-output data set from the four models is discussed. Next, steps followed in developing the higher-level universal model and optimizing its parameters are presented. Finally, conclusions are provided.

7.2: CONTEXT-SPECIFIC PRODUCTIVITY MODELS FOR CONCRETING ACTIVITY

In the following subsections, the four lower-level models, made up of fuzzy inference system models for use in predicting construction labour productivity of concreting activities under four unique construction contexts is discussed. In this study an operational definition of context for CLP modeling based on 5W1H (Who, What, Where, When, Why, and How) questions, which states that context is what constrains the four elements of a CLP model (user, model developer, model, and prevailing environment of the model) without intervening in the model development process explicitly, was adopted, as discussed in Section 5.2. Accordingly, four unique construction contexts addressing Industrial, Warehouse, High-rise, and Institutional contexts were formulated (refer to Fig. 7.1). Data on the 169 input parameters (factors and practices) and 7 process parameters (work sampling proportions) here after referred as input parameters, based on the findings of the moderated CLP system model approach discussed in Section 4.6, and output or CLP parameter was collected from each context. Accordingly, the Industrial (Context 1) had 23 input-output data instances (D1), the Warehouse (Context 2) had 16 input-output data instances (D2), the High-rise (Context 3) had 28 input-output data instances (D3), and the Institutional (Context 4) had 25 input-output data instances (D4).

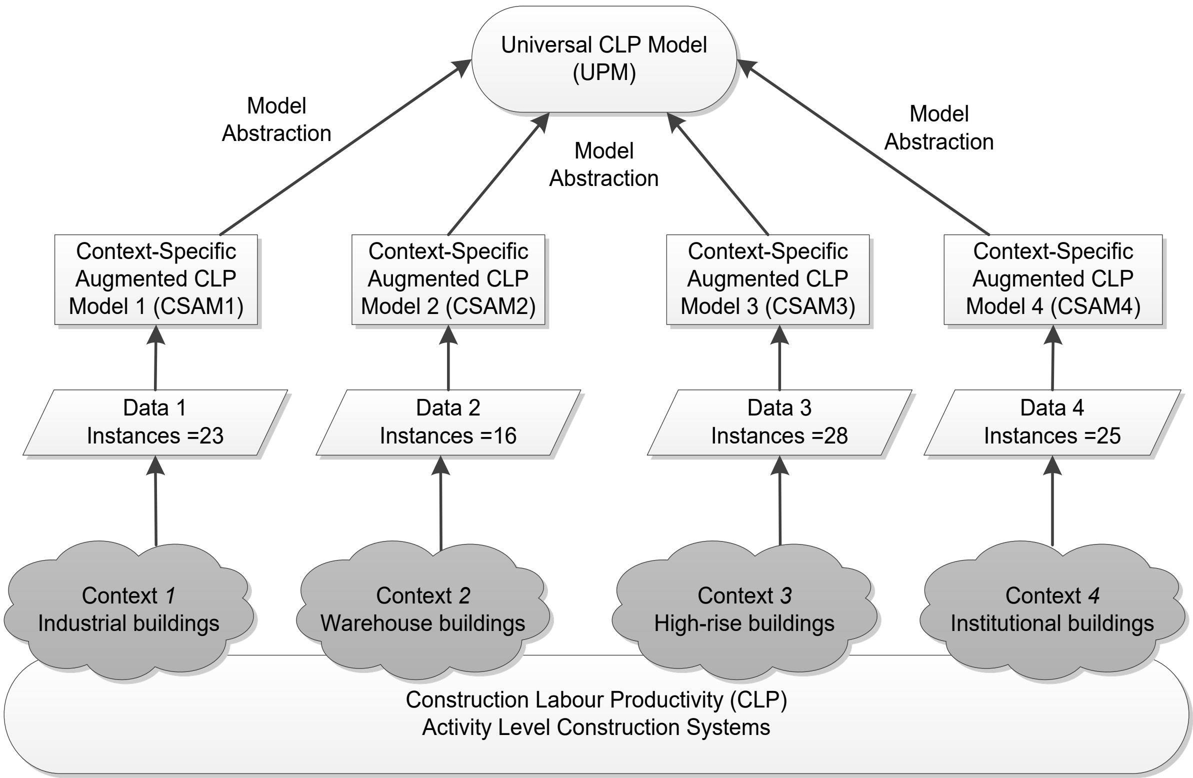

Figure 7.1: Lower-level Context-Specific CLP Models and Abstracted Universal CLP Model

7.2.1: Establishment of Universal Model Features

As discussed in Section 5.5, correlation-based feature selection process was applied for identifying the most relevant features (key input parameters), having high correlations with CLP but low inter-correlations were identified for use in development of the context-specific models. A total of 16, 7, 8, and 11 context-specific input features, representing the key influencing parameters, were selected for Context 1, Context 2, Context 3, and Context 4 CLP model developments, respectively. The selected key influencing model features for the four contexts are summarized in Table 7.1, and the full details of the model features together with their measurement scales are shown in Table 5.2.

Table 7.1: Context-Specific CLP Models: Features, Structure, and Model Parameters

Features, FIS structure, and model parameters	**Context**			
	1	**2**	**3**	**4**
	Concreting, Industrial buildings	**Concreting, Warehouse buildings**	**Concreting, High-rise buildings**	**Concreting, Institutional buildings**
Number of data instances	23	16	28	25
Number of context-specific input features	16	7	8	11
Context-specific model features	x3, x14, x45, x46, x49, x57, x106, x108, x129, x135, x138, x167, x168, y1, y2, y3, z	x45, x46, x48, x50, x108, y1, y7, z	x45, x46, x57, y1, y2, y3, y4, y6, z	x1, x5, x8, x18, x19, x46, x90, y1, y3, y4, y6, z
Combined model features	x1, x3, x5, x8, x14, x18, x19, x45, x46, x48, x49, x50, x57, x90, x106, x108, x129, x135, x138, x167, x168, y1, y2, y3, y4, y6, y7, z			
Number of augmented input features	11	20	19	15
Augmented model features	x1, x5, x8, x18,x19, x48, x50, x90, y4, y6, y7	x1, x3, x5, x8, x14, x18, x19, x49, x57, x90, x106, x129, x135, x138, x167, x168, y2, y3, y4, y6	x1, x3, x5, x8, x14, x18, x19, x48, x49, x50, x90, x106, x108, x129, x135, x138, x167, x168, y7	x3, x14, x45, x48, x49, x50, x57, x106, x108, x129, x135, x138, x167, x168, y2, y7
Fuzzification coefficient	1.5	2.5	2.0	2.0
Number of rules	6	7	6	7
Input aggregation operator	PROD	MIN	PROD	PROD
Implication method	PROD	MIN	PROD	PROD
Rule aggregation operator	MAX	PROBOR	PROBOR	PROBOR
Defuzzification method	MOM	BISECTOR	CENTROID	BISECTOR
Final accuracy (RMSE)	1.761	0.441	3.470	1.946

Note: The model features include the input features together with the output or CLP feature *z*

The four contexts, as shown in Table 7.1, have very few common features, and most features were specific to each context. The only key input variables common among the four contexts were structural element type (x46) and direct work proportion (y1). In model abstraction and universal model development, the focus is on capturing the important knowledge captured in the respective context-specific or lower-level models and the associated model features (Pedrycz and Song 2012). However, the four contexts, as shown in Table 7.1, have very few common features and most features were specific to each context. Thus, in order to capture the knowledge of each context, in terms of key influencing parameters and develop a generalized universal model, the model features from the four contexts are combined, resulting in 27 input features. The resulting 27 features represent the key factors, practices, and work sampling proportions influencing the productivity of concreting activities in any construction project, irrespective of the context, and enable the development of a context-free and general CLP model. Therefore, for each context, additional features from the combined feature list were augmented as shown in Table 7.1. Accordingly, a total of 11, 20, 19, and 15 input features were augmented to Context 1, Context 2, Context 3, and Context 4 feature lists, respectively. The augmentation process resulted in four contexts having identical input features or key influencing parameters.

Therefore, in all contexts, 27 identical input variables $\{\boldsymbol{x}_{ik}, \boldsymbol{y}_{ik}\}$ represented as $\{\vec{\boldsymbol{x}}_{ik}\}$ and a single CLP output $\{z\}$ were used. The input variables included: crew size (x1); craftsperson on-job training (x3); crew composition (x5); co-operation among craftspersons (x8); craftsperson motivation (x11); craftsperson fatigue (x12); craftsperson trust in foreman (x13); team spirit of crew (x14); level of absenteeism (x15); level of interruption and disruption (x18); fairness of work assignment (x19); job security (x23); availability of tools (x37); concrete placement technique (x45); structural element type (x46); location of work scope, distance (x48); location of work scope, elevation (x49); congestion of work area (x50); fairness in performance review of crew by foreman (x57); change of foremen (x58); site congestion (x90); project work times (x96); owner staff on site (x97); approval of shop drawings and sample materials (x101); treatment of foremen by superintendent and project manager (x106); uniformity of work rules by superintendent (x108); labour productivity measurement practice (x126); out-of-sequence inspection (x129); safety training (x135); project safety administration and reporting (x138); availability of communication devices (x143); oil price fluctuation (x167); natural gas price (x168); direct work proportion (y1); preparatory work proportion (y2); tools and equipment proportion (y3); material handling proportion (y4); travel proportion (y6); and personal proportion (y7). The list of input variables shown in Table 7.1 provide two key sources of knowledge. First, the full 27 input variables provide a complete list of the key influencing factors, practices, and work sampling proportions which have to been considered in the general planning of concreting activities, irrespective of the context (e.g. type or nature of the project). Second, if the project context is known and can be grouped under any one of the identified contexts of this study (concreting in industrial, warehouse, high-rise, or institutional context), focusing on the context-specific key input variables during planning and control stages of concreting activities will lead to improved CLP values.

7.2.2: Context-Specific Augmented CLP Models Development and Optimization

As discussed in Section 5.7 of this thesis, four original context-specific CLP models, based on the 16, 7, 8, and 11 context-specific input features shown in Table 7.1, were developed using fuzzy inference systems (FIS). FISs were composed of a family of conditional if-then rules where fuzzy sets were used in the conditions and conclusions parts of the rules. The models were developed using Fuzzy C-Means (FCM) clustering and Gaussian membership functions (MFs).The original CLP models were further optimized based on the fuzzification coefficient m in FCM clustering, membership function parameters, number of rules, and fuzzy operators and defuzzification methods. The properties of the final optimized CLP models shown in Table 5.6 and referred to as base CLP models, indicated the context-specific nature of models as the models have different number of membership functions or number of rules, fuzzification coefficients, fuzzy operators, and defuzzification methods. However, as discussed above, for the development on the universal CLP model, additional model features have been augmented to each context and the respective base context-specific models have to be expanded to accommodate the added features.

Fuzzy inference systems are highly parallel as each if-then rule is a local descriptor of the data, localized based on the fuzzy sets defined in the condition parts of the rules, and the aggregation of the if-then rules of the fuzzy inference system provides a complete description of the data (Magdalena 2015). Because of their parallel nature, fuzzy inference systems can be easily expanded by adding additional features or rules (Angelov and Buswell 2002). Taking this critical advantage of fuzzy inference models, the four context-specific models were expanded by adding the additional features. However, in order to maintain the interpretability of the augmented fuzzy inference models of each context, the numbers of rules were kept equal to that of the base context-specific model. Thus, for Industrial context (Context 1), the context-specific model was developed by adding the augmented 11 input features to the existing 16 input features. Similar to the base Context 1 model structure shown in Table 7.1, fuzzification coefficient of $m = 1.5$ was used and 6 rules were developed using FCM clustering and Gaussian membership functions. The developed augmented model was used to predict CLP values of the 23 data instances of Context 1 and the model showed an RMSE value of 1.777. Then, using the optimization framework described in Section 5.7, the membership functions of the augmented model were optimized and sensitivity analysis was carried out for the fuzzy operators and defuzzification methods. The optimized model showed an RMSE value of 1.761, which indicated 0.92% accuracy improvement over the non-optimized augmented model of Context 1. The parameters of the optimized augmented Context 1 CLP model (CSAM1) are shown in Table 7.1. Similarly, for Warehouse context (Context 2), the context-specific model was developed by adding the augmented 20 input features to the existing 7 input features. Similar to the base Context 2 model structure shown in Table 7.1, fuzzification coefficient of $m = 2.5$ was used and 7 rules were developed using FCM clustering and Gaussian membership functions. The developed augmented model was used to predict CLP values of the 16 data instances of Context 2 and the model showed an RMSE value of 0.469. Then, the membership functions of the augmented model were optimized and sensitivity analysis was carried out for the fuzzy operators and defuzzification methods. The optimized model showed an RMSE value of 0.441, which indicated 5.96% accuracy improvement over the non-optimized augmented model of Context 2. The parameters of the optimized augmented Context 2 CLP model (CSAM2) are shown in Table 7.1.

For High-rise context (Context 3), the context-specific model was developed by adding the augmented 19 input features to the existing 8 input features. Similar to the base Context 3 model structure shown in Table 7.1, fuzzification coefficient of $m = 2.0$ was used and 6 rules were developed using FCM clustering and Gaussian membership functions. The developed augmented model was used to predict CLP values of the 28 data instances of Context 3 and the model showed an RMSE value of 3.846. Then, the membership functions of the augmented model were optimized and sensitivity analysis was carried out for the fuzzy operators and defuzzification methods. The optimized model showed an RMSE value of 3.470, which indicated 9.79% accuracy improvement over the non-optimized augmented model of Context 3. The parameters of the optimized augmented Context 3 CLP model (CSAM3) having 27 input features and 6 rules are shown in Table 7.1. And, for Institutional context (Context 4), the context-specific model was developed by adding the augmented 15 input features to the existing 11 input features. Similar to the base Context 4 model structure shown in Table 7.1, fuzzification coefficient of $m = 2.0$ was used and 7 rules were developed using FCM clustering and Gaussian membership functions. The developed augmented model was used to predict CLP values of the 25 data instances of Context 4 and the model showed an RMSE value of 2.071. Then, the membership functions of the augmented model were optimized and sensitivity analysis was carried out for the fuzzy operators and defuzzification methods. The optimized model showed an RMSE value of 1.946, which indicated 6.03% accuracy improvement over the non-optimized augmented model of Context 4. The parameters of the optimized augmented Context 4 CLP model (CSAM4) are shown in Table 7.1.

A comparison of the prediction accuracies of the optimized augmented and base context-specific CLP models was then carried out using RMSE values and for each of the four contexts. For Context 1, the base context-specific model showed a better prediction of the CLP values of Context 1 with an RMSE value of 1.162, as compared to the optimized augmented CLP model (CSAM1) which has RMSE value of 1.761. However, for Context 2, the optimized augmented CLP model (CSAM2) showed a better prediction of the CLP values of Context 2 with an RMSE value of 0.441, as compared to the base context-specific model which has RMSE value of 0.467. For Context 3, the base context-specific model showed a better prediction of the CLP values of Context 3 with an RMSE value of 0.992, as compared to the optimized augmented CLP model (CSAM3) which has RMSE value of 3.470. For Context 4, the base context-specific model showed a better prediction of the CLP values of Context 4 with an RMSE value of 0.671, as compared to the optimized augmented CLP model (CSAM3) which has RMSE value of 1.946. The comparison of the RMSE results indicated the superiority of the base context-specific CLP models in predicting CLP in contexts 1, 3, and 4. However, for Context 2 the augmented CLP model showed better prediction ability, however, the improvement was marginal. Nonetheless, it should be noted that the purpose of the augmented context-specific CLP models is to facilitate the development of the universal CLP model by providing context-specific prediction of CLP using the combined input features.

7.3: Abstraction of Context-specific Augmented CLP Models

One of the objectives of this thesis is to develop a higher-level universal CLP model (UPM) from the respective lower-level models (CSAMi) as conceptualized in Fig. 7.1. This study is based on a Granular Computing and knowledge management approaches as discussed in (Zadeh 1997; Pedrycz et al. 2012; Zadeh 2008; Pedrycz 2011; Pedrycz and Song 2012; Reyes-Galaviz and Pedrycz 2015). The approach deals with an idea of knowledge transfer where a source of knowledge in terms of existing models can be used to develop an abstracted granular or universal model, which provides a better understanding of complex dynamic systems like construction (Leite et al. 2012).

Granular Computing focuses on the construction of granules and computation with granules (Yao 2007). While construction of granules deals with the formation, representation, and interpretation of granules, computation deals with the utilization of the granules in problem modeling (Yao 2007; Pedrycz

2011). Granular modeling provides a means to abstract context-specific models so that the most important relationships will used to produce a reliable prediction of CLP.

Zadeh (1997) defines an information granule as "points drawn together by indistinguishability, similarity, proximity or functionality". For example, a better representation of numeric data can be achieved via their probabilistic information granule (such as a Gaussian probability distribution) than the mean or median of the data set (Pedrycz and Song 2012). Fuzzy sets offer the important and unique feature of describing information granules of variables using linguistic terms (e.g., *large* for crew size, or *cold* for temperature) whose contributing terms will belong to a certain concept with varying degrees of membership or belongingness (Pedrycz and Gomide 2007). The essence of granular models is then to describe the association between information granules or fuzzy sets formed on the input and output spaces (Pedrycz 2011).

Information granules can be designed using either Fuzzy C-Means (FCM) clustering of the model's input-output data or by selecting representative granules from the respective models (Pedrycz and Song 2012). In case of fuzzy clustering, the primary objective is to create a family of overlapping partitions or clusters of the numerical model variable data $\{\boldsymbol{x}_k\}, k = 1,2, \ldots, N$, where N is the total number of data points or instances in the respective models. The clusters will be represented by their central element or prototype $\boldsymbol{v}$, and the degrees of belongingness captured in the form of a partition matrix U of each data point to the respective prototypes are treated as membership function. In case of selecting representative granules from the respective models, the objective is to develop higher level prototypes of the information granules by selecting a suitable subset of lower-level information granules from the respective models, where the selected subset is expected to represent all prototypes of the models to the highest extent (Pedrycz and Song 2012). The process is a combinatorial optimization problem and requires a combinatory evolutionary or population based optimization to minimize the error on reconstruction of the prototypes at the lower-level using the subset of the prototypes formed at the higher-level (Pedrycz and Song 2012).

According to Pedrycz (2011) the followings general steps are taken to realize a universal granular fuzzy model: (1) construction of information granules for the respective lower-level model input data, (2) formation of generalized information granules for input and establishing outputs from respective models, (3) development of a universal model, and (4) optimization of the universal model. The construction or formation of information granules in the first two steps, deals with the development of information granules from model input variables, and the meaningfulness of the information granules will be ensured based on the principle of justifiable granularity. Universal or granular models can then be developed using various methods: case-based reasoning, fuzzy regression, fuzzy rule-base, or fuzzy neural network (Pedrycz and Song 2012; Reyes-Galaviz and Pedrycz 2015). As discussed in the following sections, the parameters of granular models can be further optimized to improve the performance of the developed granular or universal model (Pedrycz 2011). In this study, the universal productivity model, was developed using the clustering approach on the prototypes formed from the four context-specific models, and the model utilized a granular case-based reasoning methodology (Pedrycz et al. 2012). The following steps were followed in the development of the universal productivity model:

1. Construction of lower-level information granules using FCM clustering on model input variable data (D1, D2, D3, and D4) to develop b number of prototypes.
2. Development of the generalized higher-level information granules, through clustering of the lower-level model b prototypes, to form c number of prototypes.
3. Construction of information granules of output (CLP) using an interval-based information-granule approach and the principle of justifiable granularity.
4. Development of granular output intervals based on the lower and upper bounds for use in the granular case-based reasoning approach.
5. Optimization of the generalized universal model using the area under curve values, and based on the c number of clusters built in the generalized universal fuzzy model and m-fuzzification coefficient.

In the following sub-sections, the successive design steps are discussed in detail and results and limitations are presented.

7.3.1: Construction of Lower-level Information Granules

The construction of information granules from the respective context-specific model input variables $\{\vec{\boldsymbol{x}}_{ik}\}$, where $i = 1,2,3,4$ and $k = 1,2,\dots,23$ for $i = 1$ or D1 (Industrial context), $k = 1,2,\dots,16$ for $i = 2$ or D2 (Warehouse context), $k = 1,2,\dots,28$ for $i = 3$ or D3 (High-rise context), and $k = 1,2,\dots,25$ for $i = 4$ or D4 (Institutional context), were carried out using a FCM clustering approach, which is the most commonly used method for forming granules (Pedrycz et al. 2012). The process basically classified the data set of the input variables into similar groups or clusters, each represented by a prototype. The lower-level information granules were thus made up of a collection of b prototypes $\boldsymbol{v}_{i1,}\boldsymbol{v}_{i2,}\dots,\boldsymbol{v}_{ib}$, where $i = 1,2,3,4$ and the corresponding partition matrix U relates the degree of membership in terms of the $[0,1]$ interval of each data point to the corresponding prototype. The context-specific model properties shown in Table 7.1, particularly the values of the fuzzification coefficient m and number of rules indicating the representative number of cluster centers used in the context-specific models, were used to guide the construction of the lower-level information granules. Accordingly, the number of clusters b from Context 1, Context 2, Context 3, and Context 4 datasets were set to 6, 7, 6, and 7, respectively, and the fuzzification coefficient m was set to 1.5, 2.5, 2.0, and 2.0, respectively. The subsequent collection of b number of prototypes (i.e., 6, 7, 6, and 7) from each of the four contexts resulted in a total of 26 lower-level granules.

7.3.2: Formation of Higher-level Information Granules

After forming the respective granules from the lower-level or context-specific model data, the generalized or higher-level information granules for abstracting the respective context-specific models were developed using FCM clustering on the collection of $b = 26$ prototypes. Accordingly, a second FCM clustering process was applied to the collection of the 26 lower-level granules and the resulting cluster centers or prototypes formed the higher-level information granules. The fuzzification coefficient m was set to 2.0 for optimum information granule formulation (Pedrycz and Song 2012). The resulting generalized information granules, representing the high-level information granules, were made up of a collection of c prototypes $\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,} \ldots, \dot{\boldsymbol{v}}_c$, where $\dot{\boldsymbol{v}}$ represents the higher-level prototypes.

In order to evaluate the effect of the number of clusters c on the performance of the universal model, four universal model alternatives based on *c*=3, *c*=9, *c*=15, and *c*=21 were considered. For example, Fig. 7.2, Fig. 7.3, Fig. 7.4, and Fig. 7.5 show the respective higher-level granules or prototypes $\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,}\, \dot{\boldsymbol{v}}_5, and\ \dot{\boldsymbol{v}}_9$ and the associated values of the 27 input variables for the second universal model alternative, where *c=9*. The figures display the distinct nature of each high-level granule or prototype. Conceptually each prototype represents a typical scenario in carrying out a concreting activity, where the scenario is defined using the values of the input variables which are used to predict CLP of concerting activity using the universal model.

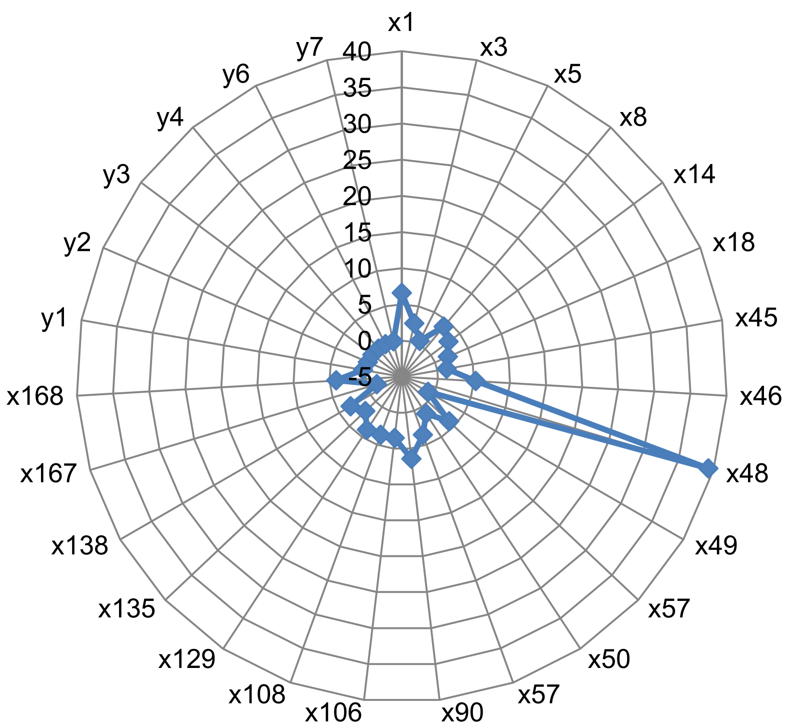

Figure 7.2: Higher-level Information Granule or Prototype $\dot{\boldsymbol{v}}_1$ for *c=9* Alternative

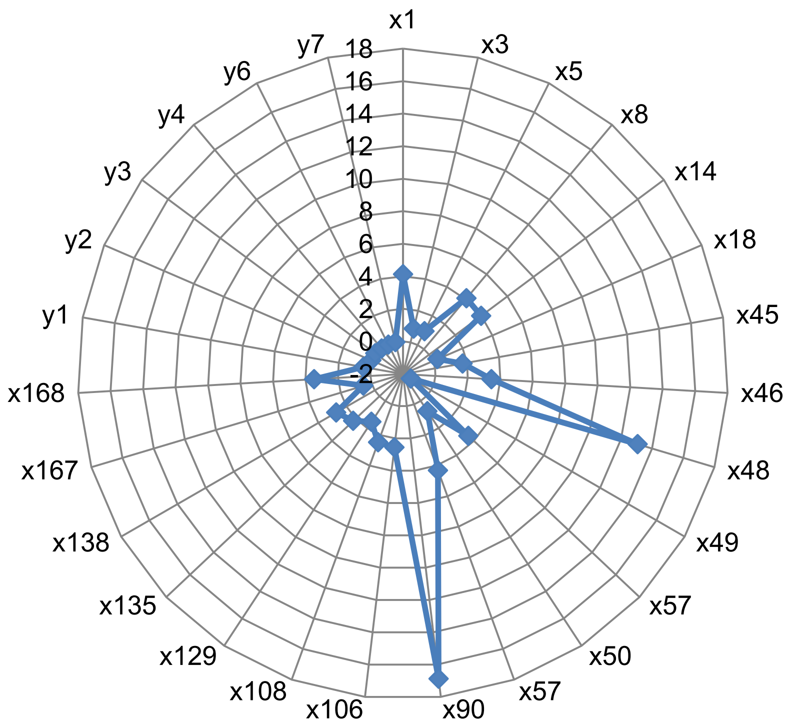

Figure 7.3: Higher-level Information Granule or Prototype $\dot{\boldsymbol{v}}_2$ for *c=9* Alternative

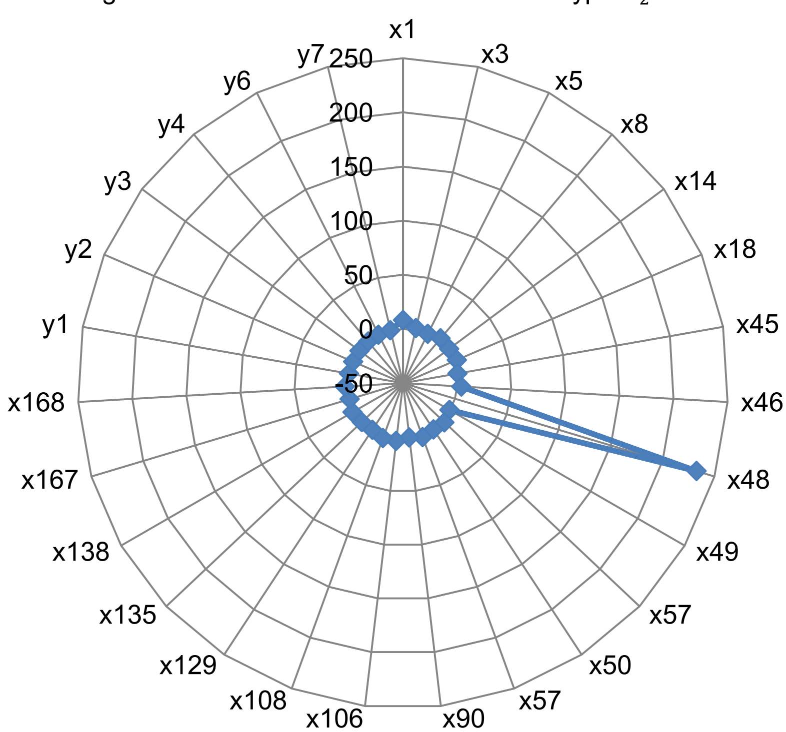

Figure 7.4: Higher-level Information Granule or Prototype $\dot{\boldsymbol{v}}_5$ for *c=9* Alternative

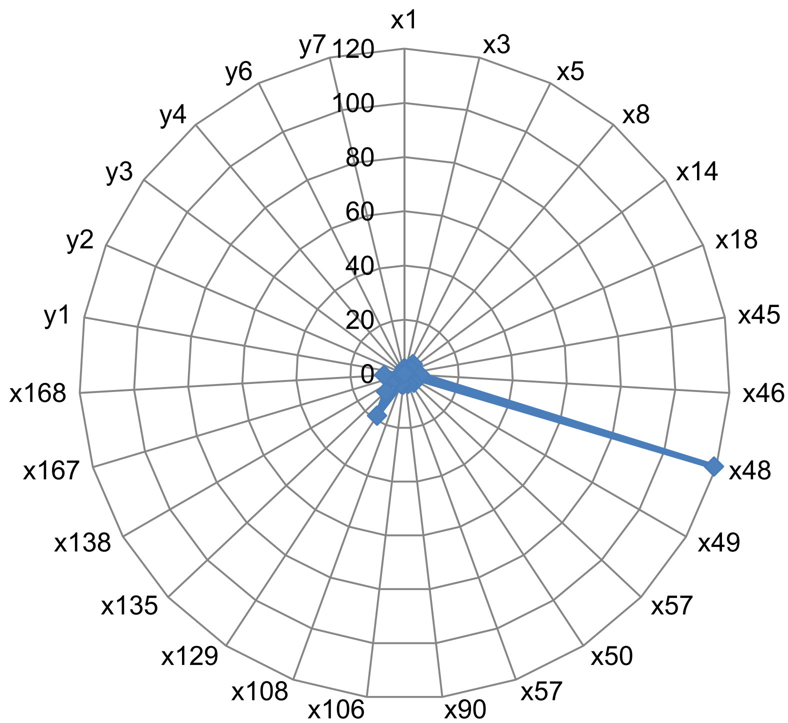

Figure 7.5: Higher-level Information Granule or Prototype $\dot{\boldsymbol{v}}_9$ for *c=9* Alternative

7.3.3: Construction of Output Information Granules

The resulting higher-level granules or prototypes, such as the ones shown in Fig. 7.2 to Fig. 7.5, represent generalized input data for predicting the output (CLP) or z. Each prototype $\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,} \dots , \dot{\boldsymbol{v}}_c$, was used as input data to the four augmented context-specific CLP models (CSAM1, CSAM2, CSAM3, and CSAM4) and four outputs were generated. The generated outputs represent the estimated CLP values $z_1, z_2, z_3, and\ z_4$ based on the knowledge stored in each of the four augmented context-specific CLP models. The four estimates of CLP were then used to develop the universal model and form the basis of the abstraction of the four models. The resulting input-output set of universal model data will thus come in the form of Eq. (7.1):

$$\begin{gathered}\dot{\boldsymbol{v}}_1 : z_{11} = CSPM1(\dot{\boldsymbol{v}}_1), z_{21} = CSPM2(\dot{\boldsymbol{v}}_1), z_{31} = CSPM3(\dot{\boldsymbol{v}}_1), z_{41} = CSPM4(\dot{\boldsymbol{v}}_1)\ ; \\ \dot{\boldsymbol{v}}_2 : z_{12} = CSPM1(\dot{\boldsymbol{v}}_2), z_{22} = CSPM2(\dot{\boldsymbol{v}}_2), z_{32} = CSPM3(\dot{\boldsymbol{v}}_2), z_{42} = CSPM4(\dot{\boldsymbol{v}}_2)\ ; \\ \cdot\ \cdot\ \cdot \\ \dot{\boldsymbol{v}}_c : z_{1c} = CSPM1(\dot{\boldsymbol{v}}_c), z_{2c} = CSPM2(\dot{\boldsymbol{v}}_c), z_{3c} = CSPM3(\dot{\boldsymbol{v}}_c), z_{4c} = CSPM4(\dot{\boldsymbol{v}}_c)\end{gathered} \tag{7.1}$$

The next step involves the determination of the universal model input-output data. The inputs for the universal models were the c prototypes ($\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,} \ldots, \dot{\boldsymbol{v}}_c$), and the output was an information granule represented using interval values Ω which was formed around the four estimated CLP values ($z_1, z_2, z_3, and\ z_4$). The output interval value forms the basis on the granulation process of the input-output data set shown in Eq. 7.1. The interval aspect of the outputs is the sought-after value of universal models, as their outputs are not specific numeric values, and will provide a range of possible CLP values for a given situation represented using the 27 combined input features. Thus, the interval based prediction of CLP provides an improvement over the commonly used single numerical estimates of CLP, as it avoids the hit or miss evaluation approach by providing a range of possible CLP values.

The construction of the output or CLP interval relied on the principle of justifiable granularity, where the formed information granules Ω are required to have sufficiently high levels of experimental evidence (coverage criterion) while maintaining high specificity (specificity criterion) (Pedrycz 2011; Pedrycz et al. 2012; Pedrycz and Song 2012; Reyes-Galaviz and Pedrycz 2015). However, the coverage and specificity requirements are conflicting as increasing the coverage aspect of the interval Ω by including sufficiently high level of the estimated CLP values ($z_1, z_2, z_3, and\ z_4$) will result into a less specific interval, thereby critically limiting the usefulness of the developed universal model. The formulation of the interval was carried out in two stages: (1) determination of the numeric representative of the four output values ($z_1, z_2, z_3, and\ z_4$) and (2) the construction of an interval bound $\Omega \in [\mathrm{e}, \mathrm{f}]$ based on its lower and upper bounds, denoted by e and f, respectively (refer to Fig.7.6).

The use of the median value as a numerical representative value is recommended (Pedrycz and Song 2012). Therefore, for a given universal model alternative, say *c=3*, three higher-level prototypes $\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,}\dot{\boldsymbol{v}}_3$ were constructed, and four output data values ($z_1, z_2, z_3, and\ z_4$) were estimated for each of the three higher-level prototype and using the four context-specific models. The four estimated output data values were then arranged in ascending order, resulting in the following list: $z_{min}, z_2, z_3, and\ z_{max}$, where z_{min} represents the lowest estimated CLP value, z_{max} represents the highest estimated CLP value, and

z_2 and z_3 represent the two middle estimated CLP values. Then, the median $Med(z)$ was computed by averaging the two middle values z_2 and z_3, as shown in Fig. 7.6.

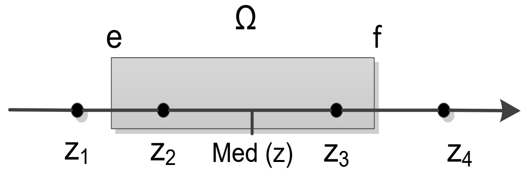

Figure 7.6: Determination of Representative Value of Output Interval

Next, the interval bounds e and f were formed around the representative value $Med(z)$ using a performance index V defined in (Eq. 7.2), where f_1 is an increasing function of cardinality of output data ($z_1, z_2, z_3, and\ z_4$) being covered by the interval Ω and is used to quantify the coverage criterion, and f_2 is a decreasing function of the support (length) that is used to quantify the specificity criterion of the interval Ω.

$$V = f_1 * f_2 \tag{7.2}$$

The objective of the process is to ensure the numeric evidence within the bounds of Ω is as high as possible (f_1) and the support of the bound is as small as possible (f_2). The following functions, where α is a positive parameter that provides flexibility in the formation of the Ω and u is interval space formed between the representative value $Med(z)$ to any one of the boundary points of granular interval (i.e., either f or e), was used (Pedrycz and Song 2012):

$$f_1 = u \tag{7.3}$$

$$f_2 = e^{-\alpha u} \tag{7.4}$$

However, as there are only four lower-level or context-specific CLP models, the output interval data will only have four values, out of which the median will be the average value of the two middle values z_2 and z_3 as shown in Fig. 7.7. Additionally, based on the value of α two granulation cases for determining the lower and upper bounds exist. In the first granulation case is shown in Fig. 7.7a, and in this case the lower and upper bounds will take the lowest and highest estimated values of CLP. Consequently, the granulation process will not provide highly specific bound $\Omega \in [e = z_{min}, f = z_{max}]$. The resulting intervals will fully fulfill the coverage requirement; however, the specificity requirement will be completely ruled out at value of α = 0, thereby making the developed universal model useless. The second granulation case is shown in Fig. 7.7b, and in this case the lower and upper bounds will take the two estimated middle values z_2 and z_3. In this case, the granulation process will provide highly specific bounds $\Omega \in [e = z_2, f = z_3]$ at the maximum α value, while coverage of the output data is achieved to a limited extent.

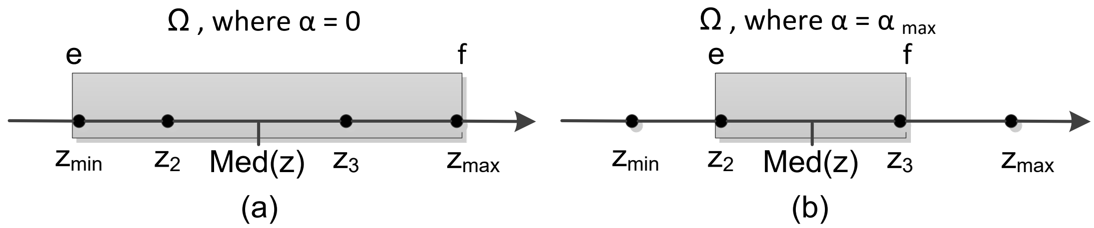

Figure 7.7: Determination of Boundary Points of Output Interval

Accordingly, the two middle values z_2 and z_3 among the four context-specific model outputs or CLP estimates $z_{min}, z_2, z_3, and\ z_{max}$ were taken to develop the lower $(e = z_2 = z^-)$ and upper bounds $(f = z_3 = z^+)$ of the intervals Ω of the universal model. The resulting input-output data of information granules will come in the following form:

$$\dot{\boldsymbol{v}}_1 \ : [z_1{}^-, z_1{}^+] \ ; \dot{\boldsymbol{v}}_2 \ : [z_2{}^-, z_2{}^+] \ ; \ . \ . \ ; \dot{\boldsymbol{v}}_c : [z_c{}^-, z_c{}^+] \tag{7.5}$$

7.3.4: Developing the Universal CLP Model

In the universal CLP model, the relationship between the higher-level information granules, represented by the c prototypes $(\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,} \dots, \dot{\boldsymbol{v}}_c)$, and their respective output intervals, represented by the respective intervals Ω of each prototype, will form the universal CLP models. Universal or granular models can be developed using various methods: case-based reasoning, fuzzy regression, fuzzy rule-base, or fuzzy neural network (Pedrycz and Song 2012; Reyes-Galaviz and Pedrycz 2015). In this study, the use of a granular case-based reasoning approach was preferred, as the universal model inputs and outputs were developed using a clustering approach, which by its very nature provides typical cases using representative prototypes. Case-based reasoning relies on specific knowledge of previously experienced cases and solves a new problem by finding a similar past case (Aamodt and Plaza 1994; Ji et al. 2012). In this study, previously experienced cases were defined based on the higher-level c prototypes and their interval-based (Ω) CLP estimates as shown in Eq. 7.5. Then the CLP estimate of a new case as defined by given input variables $\vec{\boldsymbol{x}}$ will be solved by calculating its closeness to the prototype-based cases $(\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,}\dots,\dot{\boldsymbol{v}}_{c})$. As the information granules for the input part were the higher-level information granules, represented by the c prototypes $(\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,}\dots,\dot{\boldsymbol{v}}_{c})$, each case will be represented by c number of prototypes. Thus, as discussed before, in order to evaluate the effect of the number of clusters c on the performance of the universal model, four universal models alternatives (*c*=3, *c*=9, *c*=15, and *c*=21) have been developed. For a certain input variable $\vec{\boldsymbol{x}}$, its closeness to anyone of the *c* prototypes $(\dot{\boldsymbol{v}}_{1,}\dot{\boldsymbol{v}}_{2,}\dots,\dot{\boldsymbol{v}}_{c})$ was determined using the degree of activation $u_i(\vec{\boldsymbol{x}})$ of each case or prototype. The degree of activation was computed using Eq. (7.6), where the degree of belongingness to a certain prototype $\dot{\boldsymbol{v}}_i$ is (where *m* >1):

$$u_i(\vec{\boldsymbol{x}}) = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|\vec{\boldsymbol{x}}-\dot{\boldsymbol{v}}_i\|}{\|\vec{\boldsymbol{x}}-\dot{\boldsymbol{v}}_j\|}\right)^{\frac{2}{(m-1)}}} \tag{7.6}$$

The model output in terms of lower $({z_1}^-)$ and upper bounds $({z_1}^+)$ of an interval was then determined using Eq. (7.7) and Eq. (7.8), which were derived based on the respective lower $({z_i}^-)$ and upper $({z_i}^+)$ bound values of each case or prototype:

$${z_1}^- = \sum_{i=1}^{c} u_i(\vec{\boldsymbol{x}})\ {z_i}^- \tag{7.7}$$

$${z_1}^+ = \sum_{i=1}^{c} u_i(\vec{\boldsymbol{x}})\ {z_i}^+ \tag{7.8}$$

For example, among the four universal model alternatives, two universal model alternatives are shown in Fig. 7.8 and Fig. 7.9. The first universal CLP model shown in Fig. 7.8 has 15 higher-level prototypes or *c=15* and the second universal CLP model shown in Fig. 7.9 has 21 higher-level prototypes or *c=21* and the output intervals $\Omega \in [e = z_2 = z^-, f = z_3 = z^+]$ are show for each prototype $\dot{\boldsymbol{v}}_i$.

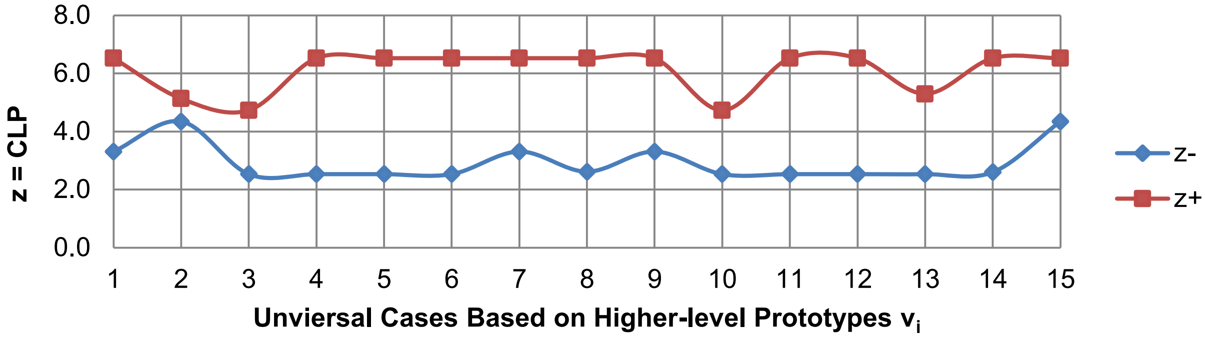

Figure 7.8: Universal CLP Model (*c*=15)

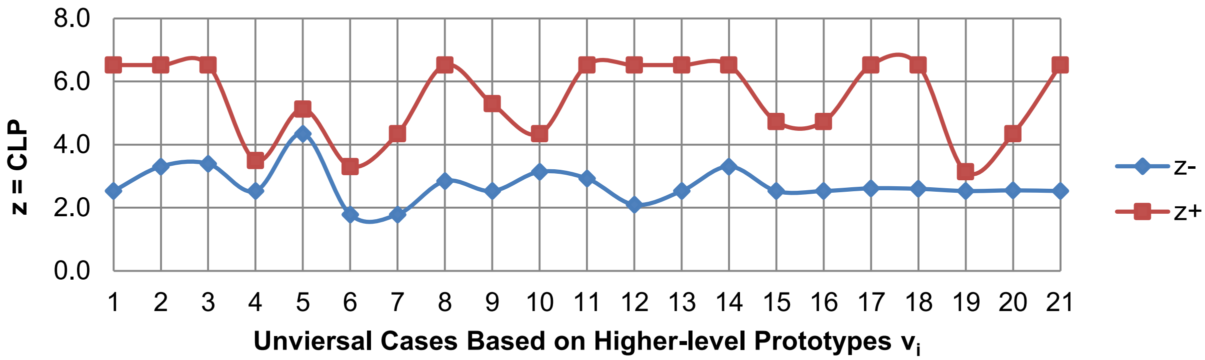

Figure 7.9: Universal CLP Model (*c*=21)

7.3.5: Optimization of the Universal CLP Model

One main advantage of universal models is that as their outputs are granular or interval based, and the interval can be optimized to improve the prediction ability of the universal models (Pedrycz and Song 2012). The optimization process was achieved using the following criteria of optimizing universal fuzzy models: *coverage criterion*, which deals with how much the interval-based CLP estimate includes the data that represents the estimates for each new case, and *specificity criterion*, which deals with to what extent the interval estimate is focused enough to provide reliable prediction of new cases.

The interval-based CLP estimates for optimizing the universal model were developed for a total of 92 cases, 23 from input-output data D1 (Industrial context), 16 from input-output data D2 (Warehouse context), 28 from input-output data D3 (High-rise context), and 25 from input-output data D4 (Institutional context). The 92 cases were randomly separated into 70% (64 cases) and 30% (28 cases) to train and test the universal model, respectively.

According to Pedrycz and Song (2012), a concise and informative characterization of universal models can be achieved using a plot of coverage of data versus specificity represented by cumulative length of the interval values. As a result, the overall quality of the universal productivity models was evaluated using the curve of coverage-specificity plots for optimizing the model parameters and selecting the one with the largest area under the curve. Following the process presented in Pedrycz and Song (2012), first the coverage requirement was quantified using a level of coverage measure. The level of coverage (*LC*) represents how much of the target output data points were included with in the universal model's interval output and, while a high value of *LC* indicates a better performing universal model, a value of *LC* close to 1 indicates an unacceptable lack of specificity of the universal model (Pedrycz and Song 2012). The level of coverage (*LC*) measure has been formulated as shown in Eq. 7.9, where N_c is the number of output or target data points covered by the universal model output, which comes in an interval form and N is the total number of cases. In this study, *N* was set to 64 for optimizing or training stage and to *N* was set to 28 for testing stage.

$$LC = \frac{N_c}{N} \tag{7.9}$$

The specificity requirement was quantified using the cumulative length (L) of the interval-based output information granules and has been formulated as shown in Eq. 7.10, where L_j is the length of an interval generated by the j^{th} data input case instance.

$$L = \sum_{j} L_j \tag{7.10}$$

Thus, for each of the training or testing input variable data points from D1, D2, D3, and D4, the degree of activation of each case was first computed using Eq. 7.6. Then, the lower and upper bounds of the universal model interval output were established using Eq. 7.7 and Eq. 7.8. Then, the cumulative length (L) was developed for the target or output data points, and the count of output data points included within the increasing intervals of L was established in order to determine the coverage requirement (LC).

However, the number of individual lower-level or context-specific models in this study was limited to four. Thus, the use of α as an optimization parameter of the universal model, which influences the determination of the interval-based output granules Ω as shown in Eq. 7.4, was not practical as the maximum α value was used in the formation of the output granules as discussed in Section 7.3.4. The parameter optimization of the universal model was thus carried out on the other two model parameters, namely, the c number of higher-level clusters built by clustering the prototypes from lower-level models, and the m fuzzification coefficient used in the determination of the degree of activation of cases $u_i(\vec{\boldsymbol{x}})$.

The universal CLP model was optimized using different combinations of the fuzzification coefficient $m \epsilon \{1.3, 1.5, 1.7, 2, 3\}$ and number of clusters of the higher-level information granules $c \epsilon \{3, 9, 15, 21\}$. It is recommended to set the fuzzification coefficient and the number of clusters to quite low and quite high values in order to compare and contrast the performance of the universal model (Reyes-Galaviz and Pedrycz 2015). The various combinations were evaluated using the area under curve values of the coverage-specificity plots, so as to explore the performance of the universal model for optimized parameters *m* and *c*. The area under curve values were determined using the ordinates of *LC* and cumulative length (L), and the maximum values of the cumulative length (*L*) leading to full coverage were considered.

First for each of the four universal model alternative (*c*=3, *c*=9, *c*=15, and *c*=21), five universal case models were developed base on the fuzzification coefficient values of *m*=1.3, *m*=1.5, *m*=1.7, *m*=2.0, and *m*=3.0. The coverage-specificity plots as shown in Fig. 7.10 were developed for optimizing the four alternative universal models using the optimization or training data cases (*N* = 64), and the coverage-specificity plot resulting in the largest area under curve were identified.

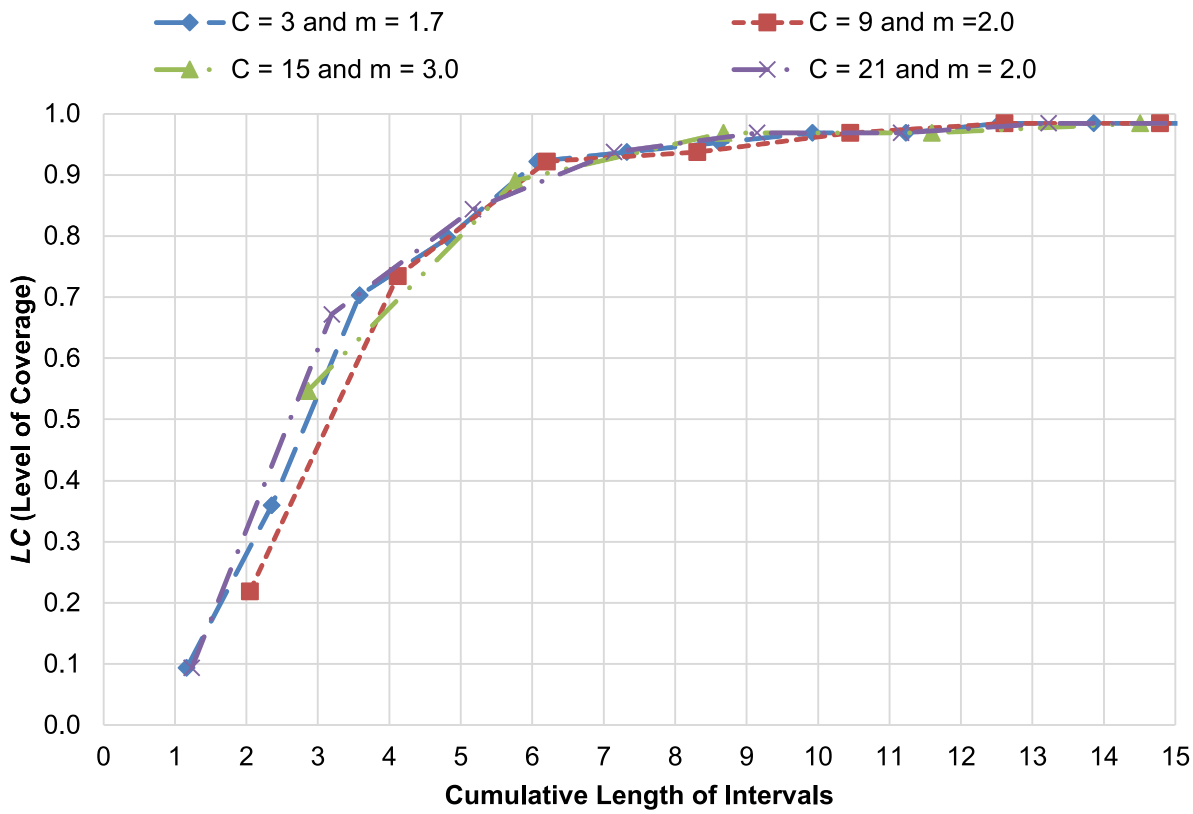

Figure 7.10: Coverage of Data *LC* versus the Cumulative Length of Interval *L* Plots

The area under curve values of the best and worst cases are summarized in Table 7.2. Accordingly, the best values of m (fuzzification coefficient) leading to the largest area under curve values for universal model (c=3), universal model (c=9), universal model (c=15), universal model (c=21) alternatives, were equal to 1.7, 2.0, 3.0, and 2.0, respectively, and the associated area under curve values were 22.35, 24.01, 23.42, and 24.43, respectively. The worst m (fuzzification coefficient) value leading to the smallest area under curve values for universal model (c=3), universal model (c=9), universal model (c=15), universal model (c=21) alternatives, were equal to 3.0, 3.0, 2.0, and 3.0, respectively, and the associated area under curve values were 21.22, 21.69, 22.55, and 22.11, respectively.

Table 7.2: Best and Worst Results of Universal CLP Model Optimization Combinations

Number of Clusters		Combination			
		Training Cases		Testing Cases	
		Best	Worst	Best	Worst
C = 3	*m*	1.7	3.0	1.7	3.0
	AUC	22.35	21.22	23.76	22.94
	LC	24.68	23.73	19.64	22.83
C = 9	*m*	2.0	3.0	2.0	3.0
	AUC	24.01	21.69	**24.26**	22.48
	LC	29.15	29.20	28.70	30.36
C = 15	*m*	3.0	2.0	3.0	2.0
	AUC	23.42	22.55	23.93	21.10
	LC	33.25	31.47	37.76	32.53
C = 21	*m*	2.0	3.0	2.0	3.0
	AUC	24.43	22.11	22.60	22.29
	LC	34.99	36.01	33.80	33.55

Note: *m* denotes the Fuzzification coefficient, AUC denotes area under curve, and *l* denote the coverage values in percent

The performance of the universal models based on the identified m values that lead to the best and worst performance measures were then evaluated using the testing data cases (N = 28). The performance results of the universal models, based on area under curve values as shown in Table 7.2, showed that the m values leading to best performance in training data cases also lead to best performance using the testing cases. Similar results were observed for m values leading to the worst performance results.

Next, each of the four universal CLP models based on *c*=3, *c*=9, *c*=15, and *c*=21 were compared to identify the optimum universal model based on the testing data cases. In Table 7.2, the best and worst results for all combinations and using training and testing data cases are shown. The area under curve based on the testing data for the four universal models based on *c*=3, *c*=9, *c*=15, and *c*=21 were found to be 23.76, 24.26, 23.93, and 22.29, respectively. The best result leading to the largest area under curve values was obtained when *m* = 1.7 and *c* = 9. Thus, the second universal model (c=9) alternative having 9 high-level prototypes or cases and based on fuzzification coefficient of $m = 1.7$ was found to be the best universal CLP model (UPM).

The universal CLP model represents a generalized context-free knowledge base and can be used to predict labour productivity of concerting activities in any context. However, the universal model has some limitations. The coverage results, computed using Eq. 7.9, and shown in Table 7.2, indicated that 28.70% of the output data points will be covered by the universal model. The low coverage value is associated with the high specificity of the developed universal model, which was caused by the limited number of individual lower-level or context-specific CLP models in study and the use of the maximum α value in the determination of the output granules. However, the high specificity of the universal model interval outputs is more important, as the interval estimates of the model will be narrow enough that a useful range of possible CLP values will be provided. On the other hand, if the coverage is high, the interval estimates of the model will be much wider, resulting in a less useful range of CLP estimate values. Another limitation arises from the low numerical accuracy of the augmented context-specific CLP models, which will hamper the performance of the universal model. Expanding the number and also improving the accuracy of the lower-level or context-specific models by examining and modeling new contexts is suggested. This will enable the optimization of the universal models using α values and lead to improved coverage performance of the universal CLP model.

7.4: CHAPTER SUMMARY

Construction labour productivity is an efficiency measure of crews in producing outputs usually at an activity level. The relationship between the parameters influencing the efficiency of crews and the achieved labour productivity is being studied using various stand-alone context-specific labour productivity models. However, the developed models suit only a specific context and most importantly, a method for generalizing the knowledge captured in the various models has not yet been fully developed. This chapter presents a granular fuzzy framework for developing universal construction labour productivity model via abstraction of context-specific CLP models. This study used four context-specific CLP models for concreting activity in order to develop a single generalized, more abstract universal CLP model. The development of the universal CLP model has been achieved using a granular case-based reasoning approach and fuzzy clustering of the prototypes developed from individual context-specific model input data points.

The resulting model has provided a granular output in terms of an interval estimate of CLP for a given value of input variables. The granular interval estimate can realistically represent CLP, which corresponds to the efficiency of a crew in converting inputs like labour, material, and equipment into project outputs. The application of the FCM clustering algorithm to develop the information granules and develop granular construction labour productivity models has been demonstrated. The performance of the universal model has fulfilled the specificity requirement, meaning that the universal model outputs will be highly focused and the interval estimate of CLP will be a narrow one, however, the universal model has shown low coverage performance. The universal CLP model will be useful for predicting labour productivity of concerting activities under any context and represents a generalized CLP knowledge base.

7.5: REFERENCES

Aamodt, A., and Plaza, E. (1994). "Case-based reasoning: foundational issues, methodological variations, and system approaches." *AI Commun.,* 7, 39–59.

Angelov, P., and Buswell, R. (2002). “Identification of evolving fuzzy rule-based models.” *IEEE T. Fuzzy Syst.*, 10(5), 667–677.

Dick, S., and Kandel, A. (2001). "Granular computing in neural networks." *Granular Computing: An Emerging Paradigm*, New York: Physica-Verlag Heidelberg, US, 275–305.

Ezeldin, A. S., and Sharara, L. M. (2006). "Neural networks for estimating the productivity of concreting activities." *J. Constr. Eng. Manage*., 132(6), 650–656.

Fayek, A. Robinson, and Tsehayae, A. A. (2012). "Modeling construction labour productivity using fuzzy logic and exploring the use of fuzzy hybrid techniques." *Fuzzy Information Processing Society (NAFIPS), 2012 Annual Meeting of the North American,* IEEE, Berkeley, California, US, 1–6.

Handa, V., and Abdalla, O. (1989). “Forecasting productivity by work sampling.” *Constr. Manage. Econ.*, 7(1), 19–28.

Hanna, A. S., Taylor, C. S., and Sullivan, K. T. (2005). "Impact of extended overtime on construction labor productivity." *J. Constr. Eng. Manage*.,131(6), 734–739.

Hanna, A., Russel, J., Gotzion, T., and Nordheim, E. (1999a). “Impact of change orders on labor efficiency for mechanical construction.” *J.Constr. Eng. Manage.*, 125(3), 176–184.

Hanna, A., Russel, J., Nordheim, E., and Bruggink, M. (1999b). “Impact of change orders on labor efficiency for electrical construction.” *J.Constr. Eng. Manage.*, 125(4), 224–232.

Hendrickson, C., Martinelli, D., and Rehak,D. (1987). "Hierarchical rule-based activity duration estimation." *J. Constr. Eng. Manage*.,113(2), 288–301.

Horman, M. J., and Kenley, R. (2005). "Quantifying levels of wasted time in construction with meta-analysis." *J. Constr. Eng. Manage*.,131(1), 52–61.

Ibbs, C. (1997). “Quantitative impacts of project change: Size issues.” *J. Constr. Eng. Manage.*, 123(3), 308–311.

Ji, S., Park, M., and Lee, H. (2012). "Case adaptation method of case-based reasoning for construction cost estimation in Korea." *J. Constr. Eng. Manage*.,138(1), 43–52.

Leite, D., Ballini, R., Costa, P. and Gomide, F. (2012). “Evolving fuzzy granular modeling from nonstationary fuzzy data streams.” *Evolving Sys.*, *3*(2), 65–79.

Magdalena, L. (2015). "Fuzzy rule-based systems." *Springer Handbook of Computational Intelligence*. Springer Berlin Heidelberg, 203-218.

Mosehli, O., Assem, I., and El-Rayes, K. (2005). “Change orders impact on labor productivity.” *J. Constr. Eng. Manage.,* 131(3), 354–359.

Moselhi, O., Charles, L., and Fazio, P. (1991). “Impact of change orders on construction productivity.” *Can. J. Civ. Eng.*, 18(4), 484–492.

Pedrycz, W. (2011). "From fuzzy models to granular fuzzy models." *Fuzzy Logic and Applications*: *Proc., 9th International Workshop, WILF,* Trani, Italy, 75–82.

Pedrycz, W., and Gomide, F. (2007). *Fuzzy systems engineering: Toward human-centric computing*. John Wiley and Sons, Hoboken, New Jersey, US.

Pedrycz, W., Russo, B., and Succi, G. (2012). "Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity." *Appl. Soft Comput.,* 12(8), 1985–1995.

Pedrycz, W., and Song, M. (2012). “Granular fuzzy models: A study in knowledge management in fuzzy modeling.” *Intern. J. Approx. Reasoning*, 53(7), 1061–1079.

Portas, J., and AbouRizk, S. (1997). "Neural network model for estimating construction productivity." *J. Constr. Eng. Manage*., 123(4), 399–410.

Reyes-Galaviz, O. F., and Pedrycz, W. (2015). “Granular fuzzy models: Analysis, design, and evaluation.” *Intern. J. Approx. Reasoning*, 64, 1–19.

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). "Modeling construction labor productivity." *J. Constr. Eng. Manage*., 116(4), 705–726.

Tsehayae, A. A., and Fayek, A. Robinson. (2014). "Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects." *Can. J. Civ. Eng.,* 41(10), 878–891.

Tsehayae, A. A., Pedrycz, W., and Fayek, A. Robinson (2013). "Application of granular fuzzy modeling for abstracting labour productivity knowledge bases." *Proc., IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint Congress,* IEEE, Edmonton, Alberta, Canada, 1096–1101.

Yao, J. (2007). "A ten-year review of granular computing." *Proc., Intl. Conf. on Granular Computing,* IEEE, Silicon Valley, California, US, 734–739.

Yi, W., and Chan, A. (2014). "Critical review of labor productivity research in construction journals." *J. Manage. Eng.,* 30(2), 214–225.

Zadeh, L.A. (1997). "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic." *Fuzzy Sets and Systems,* 90, 111–117.

Zadeh, L. A. (2008). "Is there a need for fuzzy logic?" *Inf. Sci.,* 178(13), 2751–2779.

Chapter 8: Conclusions and Recommendations[8]

This chapter presents a review of the work conducted in this research and summarizes the contributions. Limitations of the research and the developed context-specific and universal construction labour productivity models are discussed, and recommendations for future research are outlined.

8.1: RESEARCH SUMMARY

Construction labour productivity (CLP) is an efficiency measure of an activity-level open system that deals with the process of converting inputs (material, information, etc.) to outputs (project components) using labour as the chief transformation mechanism. Labour productivity is of critical importance to the construction industry, as it directly affects the profitability and competitiveness of construction companies, and it is therefore a frequently researched topic. Nevertheless, labour productivity continues to be a major source of construction risk and exhibits the highest variability among construction resources. The construction industry is constantly searching for ways to improve labour productivity, but the industry has lacked crew-level CLP models capable of explaining which parameters cause productivity to change and by how much. Construction labour productivity is affected by numerous context-sensitive subjective and objective influencing parameters. Modeling CLP remains a challenge, since for a given context, the complex impact of the multiple parameters have to be considered simultaneously, without sacrificing accuracy and interpretability of developed context-specific CLP models. Previous studies also had limitations in proving the ability of work sampling proportions to explain the variability of CLP. Additionally, previous CLP models lacked a clear and explicit representation of context, and an approach for transferring the knowledge represented in the context-specific CLP models from one context to another was missing. Furthermore, a universal model that represents versatile knowledge that can be used in any context was needed.

---

[8] Parts of this chapter have been published in Canadian Journal of Civil Engineering, Volume 41, Issue 10, pp. 878-891; submitted for publication in Journal of Construction Innovation: Information, Process, Management, JCI, 36 manuscript pages, submitted July 28, 2015; and submitted for publication in Journal of Construction Engineering and Management, JCEM, 33 manuscript pages, submitted June 30, 2015.

The main aim of this thesis is thus to present a structured methodology for the development of interpretable and accurate context-specific and universal CLP models that facilitate a better understanding of the parameters that influence CLP. The methodology examined the effect of the numerous context-sensitive influencing system parameters, made up of input parameters (subjective and objective factors and practices), process parameters (work sampling proportions), causing the complex variability of the output parameter (CLP) using data-driven and optimized context-specific fuzzy inference system models and granular case based universal CLP model.

The research in this thesis was conducted mainly in four main stages: (1) identifying, classifying, quantifying, and documenting the input and process parameters influencing CLP, (2) identifying key input parameters (factors and practices) influencing labour productivity using expert and data-driven approaches, (3) developing a system-based labour productivity modeling framework for establishing the role of work sampling proportions, in addition to the input parameters, in labour productivity modeling, and (4) studying the effect of context in CLP by developing and optimizing context-specific CLP models and formulating context adaptation and context abstraction frameworks.

8.1.1: The First Stage

The second chapter reviews existing CLP modeling literature, identifies limitations of past CLP studies, and presents the detailed description of the research methodology of this study. The methodology used for identifying, classifying, and quantifying the input, process, and output system parameters was described. A hierarchal list made up 169 input parameters and seven work sampling proportion influencing CLP was developed. Then, the details of the data collection methodology were presented together with the details of the custom-made, server-based database tool, called ProductivityTracker©. Additionally, the research ethics procedure developed to ensure the anonymity and confidentiality of research participants was summarized. Using several data collection techniques (factor survey, factors and practices documentation, work sampling studies, foreman delay surveys, craftsman questionnaires, and productivity measurements) the system parameters were documented from 11 projects across Alberta, Canada. However, adequate data was available only for the following trades: concreting, electrical, mechanical, and boilermakers.

8.1.2: The Second Stage

The large input parameters feature space, made up of the influencing input parameters (factors and practices), had to be reduced to maintain the interpretability and accuracy of the CLP models. In the third chapter, a methodology for identifying the key input parameters critically influencing labour productivity based on a hybrid expert and data-driven approach was developed and implemented. A methodology based on factor surveys for collecting expert opinions from different contexts was developed using two survey forms, namely the project management survey (PM survey) and the trade survey.

A total of seven companies involved in commercial (three), residential (two), and industrial (two) projects participated in the factor survey study and provided a total of six projects. The projects under study have been divided into two main context categories based on industry type. The first deals with the building construction context and involves commercial and residential projects; the second deals with the industrial construction context and involves industrial plant shutdown projects. However, since there was only one project in the commercial category, the survey data from that project were merged with those of the residential projects and the context was classified as the building project context. Based on the survey responses from project manager and trade level project participants under the building and industrial contexts, the key input parameter influencing CLP positively and negatively was identified. The internal consistency of the survey responses were examined using Cronbach’s alpha values, and using statistical analysis the difference in perspective between contexts and respondents were examined. Using the expert-driven approach, the following tasks were accomplished: (1) verifying the hierarchal list of parameters influencing CLP; (2) establishing the properties of the input parameters by identifying their level of existence together with the most appropriate project member to be targeted, so as to gather accurate parameter values; and (3) establishing the context-specific nature of key parameters influencing

CLP. Based on the findings of the expert-driven approach, data were collected on the input parameters, and key parameters were identified using the data-driven approach.

The data-driven approach relied on the field data for construction labour productivity and the influencing parameters; therefore, the approach is applicable to identification of parameters at activity level, where labour productivity values are studied, and provided the much needed key parameters for CLP model development. However, the analysis focused on concreting data collected from six building projects, as the data instances from the six structural elements were combined, which, compared to the other activities studied, produced the largest data set with a total of 92 data instances. The data-driven methodology was based on feature selection technique, and the most critical parameters were identified using field data collected for each of the hierarchal parameters. Feature selection was carried out using correlation-based feature selection algorithm.

8.1.3: The Third Stage

In the fourth chapter, the system-based labour productivity modeling framework for establishing the role of work sampling proportions in addition to the input parameters in labour productivity modeling was presented. The role of process variables or work sampling proportions in CLP modeling was formulated using the collected field data. The fundamental assumption of activity models—that CLP improves if more time is spent on direct work activities—was tested using scatter plots, correlation analysis, and linear and nonlinear regression analysis. The validity of the activity models investigated the relationship between CLP and seven work sampling proportions for eight activities: concreting for columns, slabs, and walls; electrical box installation, piping, and wire pulling; and shutdown overlays and shield installations. A stepwise multivariate linear regression analysis was also carried out between the process and output or CLP variables to examine the capability of process variables in providing a credible explanation to the variability of CLP, and based on the results of the null hypothesis tests, inferences on the role of process variables in CLP modeling were made. The investigation showed that direct work proportions are not significantly correlated to CLP and accurate prediction of CLP is not possible with either linear or nonlinear regression models. No significant correlations between the proportion of direct work and CLP could be observed, so direct work proportions cannot be used as surrogate measures of CLP; furthermore, activity models based on other work sampling proportions are not able to explain the variability of CLP and need additional explanatory parameters to improve their predictive capability.

Accordingly, a system-based modeling framework was proposed, developed, and tested for concreting (concrete placement) activity. The system model consisted of input (*I*), or independent, variables representing key parameters (factors and practices) influencing CLP; process (*P*) variables representing the seven work sampling proportions; and an output (*O*), or dependent, variable representing CLP. In the system-based modeling framework, depending on the mediation or moderation role of the *P* variables in explaining the variability of CLP, three different paths were considered. The first path was based on the *I*–*O* relationship and comprises the factor CLP model. The second path was based on the *I*–*P*–*O* relationship and assumes that process variables have a mediating effect; it comprised a "mediated system" CLP model. In the mediated system CLP model, the assumption was based on complete mediation, where the *I* variables influence the *P* variables as mediator variables, which in turn influence the output or dependent variable (*O*). The third path was based on the (*I* and *P*)–*O* relationship and assumes that *P* variables have a moderating effect; it comprised a "moderated system" CLP model. In the moderated system CLP model, the assumption was that the *P* variables, as moderator variables, affect the direction and strength of the relationship between the *I* and *O* variables. The mediation and moderation effect of the process variables were tested by developing artificial intelligence technique-based models, namely, neural networks, Mamdani fuzzy inference systems, and Sugeno fuzzy inference systems, and evaluating which model and path provided the most accurate results. Timeliness, precision, repeatability, and accuracy performance metrics were used to determine the overall accuracy of a given model path, and the three model paths were tested using field data collected for this research.

The analysis results showed that the moderated system model was the most accurate; it also had the best performance for timeliness and precision measures. The moderated system model proved that process variables have a moderating effect on factor models. The analysis results further indicated: (1) a causal relationship between parameters and CLP exists, and optimization of CLP can be achieved by adjusting the parameters or key factors and practices influencing it; and (2) work sampling proportions can be integrated with parameters to provide a better prediction of CLP. However, the system modeling analysis was carried out for concerting activity only, as the other activities (electrical and shutdown) did not have adequate data instance for system analysis.

8.1.4: The Fourth Stage

Based on the developed system-based CLP modeling framework, a series of context-specific and a universal CLP model were developed for concreting activity. The fifth chapter presents an operational definition of context for CLP modeling, and the associated context attributes were developed based on the 5W1H (Who, What, Where, When, Why, and How) question and answers approach. In this study, the context attributes were assumed to be static for any given project under study.Then, a framework for the development of context-specific CLP models based on fuzzy inference systems (FIS) was developed. The framework first formulated the unique contexts of the studied construction projects and for concreting activity, using 5W1H approach and the projects sharing similar contexts were combined, resulting in four unique construction contexts. Thus, four original context-specific CLP models addressing in Industrial, Warehouse, High-rise, and Institutional contexts were developed. Additionally a generic CLP model, based on the combined data set of the unique contexts, was developed. Finally, the learning ability of the developed FIS CLP models was improved using a multi-objective optimization framework which optimizes the following model parameters: (1) the fuzzification coefficient m in FCM clustering, (2) membership function parameters, (3) number of rules, and (4) fuzzy operators and defuzzification methods. As optimizing the numerous model parameters at once will create a large search space, a step by step optimization process was applied, based on the assumption that the process will create a smaller search space where the optimization process has a better chance of arriving at optimal model parameter values.

The resulting base context-specific models were validated using leave-one-out validation strategy and the generic CLP model was validated using 70%–30% split of data into training and testing subsets.

The context adaptation framework for transferring knowledge among contexts was then developed. In the sixth chapter, the procedure for linear and nonlinear adaptation of context-specific CLP models was formulated based on the transformation of the membership functions and further sensitivity analysis of adapted models to various fuzzy operators and defuzzification methods. Using the four context-specific CLP models for concreting activity, the developed context adaptation framework was implemented. The results indicated that in terms of model accuracy, none of the adapted models performed better than the base models of a given context; however, this is expected as the base context models were developed and further optimized using the context-specific key variables influencing CLP and the associated dataset. However, the comparison of the adapted models' accuracy with the accuracy of the original context-specific models (models before optimization) showed promising results for the Industrial and Warehouse contexts; thus, considering the effort required for collecting data on all influencing variables and developing and optimizing new models, the application of the context adaptation framework can provide a simpler and efficient model development alternative for Industrial and Warehouse contexts. The adapted models will be useful for predicting CLP in the early stages of project planning. Additionally, the sensitivity analysis on fuzzy operators and defuzzification methods did not show significant improvement of the adapted model accuracy. The adapted context-specific models were validated using modified Willmott agreement index between the adapted and base models of a given context.

Finally, a framework for the development of the universal CLP model was proposed and tested. The framework abstracted the four context-specific fuzzy models into a single generalized, more abstract universal CLP model. The universal model was developed using Fuzzy C-Means clustering of the four context-specific data sets and a granular case-based reasoning approach. The universal model has provided a granular output in terms of an interval estimate of CLP for a given value of input variables, and the granular interval estimate can realistically represent CLP. The performance of the universal model has fulfilled the specificity requirement, meaning that the universal model outputs will be highly focused and the interval estimate of CLP will be more useful. The universal CLP model was validated using 70%–30% split of data into training and testing subsets. However, the universal model faces the following limitations: (1) significant data demand as the model requires field data on a total of 27 key influencing parameters, and gathering extensive field data for this many parameters will be time consuming and expensive, (2) the number of context-specific models used for abstraction was limited and additional context-specific models are needed to expand the knowledge base of the universal model, and (3) the context-specific models used for abstraction also had low accuracy in predicting CLP, which in turn limited the accuracy of the developed universal model.

In the development and optimization of the several context-specific base and adapted models, and the universal CLP model, particular emphasis has been made on the interpretability aspect of the models. The number of rules and the number and overlap of the membership functions in the context specific CLP models, and the number of cases or cluster centers in the universal CLP model were designed in such a way that the developed models will have adequate transparency, so that users can easily comprehend the model and understand the relationships captured by it.

8.2: RESEARCH CONTRIBUTIONS

This research described in this thesis presents several frameworks that are relevant for researchers and industry practitioners. The details of the academic contributions, relevant to academic researchers, and industrial contributions, relevant to industry practitioners, are presented in the following sub-sections.

8.2.1: Academic Contributions

The main academic contributions of this research are summarized as follows:

- *Development of a detailed methodology for quantifying subjective and objective parameters (factor, practices, and work sampling proportions) influencing CLP*: The developed parameter quantification methodology explicitly defines each parameter, establishes measurement scales, and formulates data collection cycles and data sources. Hence, the parameters quantification methodology developed in this study provides researchers a means for gathering accurate data on not only objective factors, but also subjective factors and practices influencing CLP.

- *Development of a hybrid expert and data-driven methodology for identification of the context-specific enablers as well as barriers to better CLP*: The proposed methodology evaluates and ranks the influencing parameters (factors and practices) based on their positive or negative influence on CLP, and for the first time in CLP research, the approach combines expert and data driven approaches to identifying the key context-specific parameters influencing CLP. In this two staged methodology, expert knowledge is elicited using context-centered surveys so as to verify the identified list of parameters and establish the properties of the parameters in terms of existence or frequency and context sensitivity of the parameters, and the data-driven approach, employing a feature selection algorithm, is used to identify the actual key influencing parameters using field data.
- *Evaluating the usefulness of relying on work sampling proportions like direct work or tool time in CLP modeling:* Previous construction labour productivity studies had limitations in proving the ability of work sampling proportions to explain the variability of CLP. This study tested the assumption that direct work proportions are highly correlated to CLP and can provide reliable predictions of CLP using field data consistently collected from various projects and using accurate measures of CLP, standard WS categories, a strict data collection protocol, and a wide variety of activities ranging from commercial concreting of structural elements to boiler maintenance work in shutdowns. The results showed the direct work proportions were not significantly correlated to CLP and the limitations of work sampling proportions in accurately predicting CLP were formulated.
- *Development of a novel system model framework for CLP modeling*: The proposed framework provides an evaluation of how input parameters (factors and practices) and process parameters (work sampling proportion) influence CLP. The framework provided a better understanding of CLP, the

parameters influencing CLP, and how work-study methods like work sampling can be integrated to provide accurate CLP prediction and analysis tools.

- *Development of a framework for developing and optimizing interpretable and accurate context-specific CLP models:* The proposed framework addresses the challenges faced in past CLP models by developing interpretable and accurate context-specific fuzzy inference CLP models based on a clear and explicit representation of context so as to explain the impact of multiple context-specific subjective and objective parameters on CLP, while requiring limited data for development. The framework also advances the state of art of fuzzy hybrid modeling using a multi-objective optimization framework for improving not only the accuracy, but also the interpretability, of developed fuzzy inference CLP models.
- *Development of a novel context adaptation framework for adapting CLP models from one context to another*: The proposed framework enables the transfer of knowledge among existing fuzzy inference based productivity models. The framework was based on linear and nonlinear adaptation of the membership functions of the context-specific CLP models, followed by evaluation of the adapted models using combination of fuzzy operators and defuzzification methods. The context adaptation framework enables the reuse of existing CLP models and provides a simpler and efficient alternative for developing CLP models.
- *Development of a novel framework for developing and optimizing universal CLP models*: The proposed framework enables the abstraction of the knowledge bases represented in existing context-specific productivity models so as to develop a single generalized and more abstract universal CLP model. The framework, which was based on Fuzzy C-Means clustering and granular case-based reasoning approaches, provides a realistic interval based estimate of CLP.

8.2.2: Industrial Contributions

The main industrial contributions of this research are summarized as follows:

- *Development of key parameters made up of factors, practices, and work sampling proportions influencing CLP*: The study presents a general multilevel list of parameters (factors, practices, and work sampling proportions) affecting construction labour productivity; and identifies the key parameters influencing CLP in various contexts. The identified key parameters provide industry practitioners a useful insight on the issues to focus on during construction planning and execution phases so as to improve labour productivity and project profitability.
- *Development of a structured construction labour productivity data collection methodology and data collection protocol*: One of the main challenges in studying and improving construction labour productivity is the lack of reliable data. This study developed a detailed and structured data collection methodology for collecting data on the numerous influencing parameters from various construction projects using the following methods: factor surveys, factors and practices documentation, work sampling studies, foreman delay surveys, craftsman questionnaires, and productivity measurements. A comprehensive data collection protocol to ensure the quality and consistency of data was also developed. The data collection protocol provides detailed guidelines for industry practitioners interested in carrying out labour productivity improvement studies.
- *ProductivityTracker© tool for advanced data storage and analysis*: The developed tool has a setup, data inventory, and report/analysis modules, and can store and analyze a variety of construction productivity variables data collected from numerous construction projects. The tool provides construction companies a means to store and analyze vast amounts of construction productivity related data and facilitate the development of construction productivity improvement strategies. The tool's security setting also encourages the implementation of company or industry wide productivity improvement studies, as users are provided with different levels of access, and data anonymity and security are preserved.

- *Development of a series of context-specific and a universal CLP models*: In this study, a number of context specific CLP models were developed for predicting labour productivity of concreting activities under four unique settings related to Industrial, Warehouse, High-rise, and Institutional construction contexts. The study has also developed the first of its kind generalized and context-free universal CLP model for concreting activities; however, the model had low coverage performance. Industry practitioners can use the models developed through this study to predict and analyze CLP of concreting activities for new projects, either using context-specific CLP models in cases where a given new project's context attributes based on 5W1H (Who, What, Where, When, Why, and How) questions resemble any one of the studied contexts, or using the universal CLP model in case of projects having completely new contexts.

The findings and developed models this study can be applied by Industry practitioners (project managers, planners, supervisors, etc.) in future projects. Accordingly, the following four potential areas of application of the developed context-specific and universal concreting activity CLP models are identified:

- *Estimating CLP for use in construction project cost estimation and scheduling*: The developed models can be used provide reliable prediction of CLP values for concreting activity performed under industrial, warehouse, high-rise, or institutional contexts. A general interval based CLP estimate can also be predicted for planning concreting activities in future projects. The prediction will be made using values of the key influencing parameters (e.g. crew size, crew composition, availability of tools, location of work scope, direct work proportions, etc.), included in the respective context-specific and universal CLP models.
- *Assisting structured CLP focused scenario analysis*: Using the developed CLP models, industry practitioners can carry our scenario analysis and examine the impact of one or more key influencing parameters (e.g. concrete placement technique or site congestion) on CLP. Scenario analysis will involve changing the level of any one of the influencing parameters, while maintaining the level of the other influencing parameters at expected values of a given scenario, and reviewing the associated

increase or decrease of CLP values, which can be easily carried out using the developed CLP models and for various contexts. The correlation coefficients of the key influencing parameters included in the CLP models indicate the positive and negative effect of respective parameters, and the associated fuzzy if-then rules will translate the impact of the change of the influencing parameters on the output (CLP). As such, the individual effect of the parameters can be evaluated and potential improvements in CLP can be generated using such scenario analysis.

- *Facilitating the training of project supervisors*: The identified key influencing parameters made up of factors, practice, and work sampling proportions can be included in training manuals so as to improve the knowledge of project supervisors leading to better planning and control of construction projects. Additionally, the developed CLP models provide a virtual representation of the construction conversion process, and can be used as part of productivity training programs, so as to provide detail description of the construction process and how the influencing parameters can be used to better understand the efficiency of the process.
- *Facilitating the adoption of best practices*: In past CLP studies a number of best practices have been proposed. However, substantiating the possible gain in CLP due to the adoption of such best practices remains difficult. This study has developed a number of CLP models, which can quantify the expected gains in CLP due to the adoption of best practices such as implementing labour productivity measurement practices or safety training, and the predicted gains in CLP can be further examined using case study projects.

8.3: Research Limitations and Recommendations for Future Research and Development

This research provides a structured methodology for the development, optimization, adaptation, and abstraction of interpretable and accurate context-specific CLP models that facilitates a better understanding of the parameters that influence CLP. The different frameworks that make up this study's methodology examined the effect of the numerous context-sensitive CLP influencing system parameters using a series of context-specific and generalized universal CLP models. Despite achieving its initial intended aims, the research has certain limitations. The limitations and recommendations for future research, so as to address the current limitations, are provided in the following subsections.

8.3.1: CLP Model Scope

This study carried out extensive data collection over a 29-month period and collected data from 11 projects across Alberta, Canada, where four trades were studied: concreting, electrical, mechanical, and boilermakers. However, limited data were available for the mechanical trade. Also, only the concreting activities had enough data instances to carry out the system based modeling approach, and had adequate number of projects for the context focused modeling investigations, which address context-specific model development and optimization, context adaptation, and context abstraction frameworks. Further studies are required to collect additional data instances for concreting activity so as to improve the developed models' accuracy. Also, additional research is required to expand the number of context-specific models for the other trades (electrical, mechanical, and boilermakers) by examining and modeling new contexts. Additional investigation with other labour-intensive activities, such as welding, piping, and scaffolding is recommended to further enhance the developed CLP modeling frameworks. As the physical measurement of outputs for certain activities (e.g. piping in industrial projects) could prove to be difficult, the use of performance factor (PF) for measuring CLP is recommend in studying such activities. Furthermore, as construction is a physically demanding occupation and requires the handling of heavy loads and carrying out repetitive tasks and in this study the impact of such works tasks (e.g. fatigue) have been studied by accounting for overtime and documenting work related injuries. However, additional study on ergonomic analysis of work tasks is recommended in order to evaluate the relationship between work tasks and the physical and cognitive capabilities of construction workers, as such analyses could lead to improved capability of construction workers and lead to better CLP (Pinto et al. 2011).

8.3.2: CLP Model Development and Optimization

In this study, since the presence of many influencing parameters will create a curse of dimensionality and reduce the interpretability and usefulness of the developed fuzzy models, the feature space was reduced using a correlation based feature selection approach, and key influencing parameters were identified for development of CLP models. In future CLP studies, the use of domain knowledge and the findings of this and similar researches identifying key influencing parameters is recommended to reduce the feature space, and facilitate the development of accurate and interpretable CLP models. Additional study on data instance selection is also needed so as to review the effect of the individual data instances on the CLP models. The use of bi-clustering algorithms, which can simultaneously cluster the data instances and features space matrix and find useful patterns between the influencing parameters (i.e., features) and CLP, is recommend for further investigation. In bi-clustering or co-clustering, the feature space made up of D data instances and F features is classified in such a way that features of the class F_k are responsible for creating the class of data instances D_k, and resulting in pair (D_k, F_k) called a bi-cluster (Madeira et al. 2004). Bi-clustering has successfully applied for dimensionality reduction in biomedicine, text mining, and marking fields, and could provide improved results in CLP analysis (Busygin et al. 2008).

Additionally, the fuzzy rule-bases of the context-specific CLP models were developed using a Fuzzy C-Means (FCM) clustering algorithm, which is the most commonly employed data-driven approach for the development of fuzzy if-then rules. However, FCM clustering approach partitions the combined input-output variable space by assuming that the input and output variables have equal importance or weight. Therefore, further improvement of the FCM clustering algorithm is needed, as more emphasis or weight should be given to the output variable space. Thus, an improved clustering algorithm based on conditional clustering is recommended for further investigation. Furthermore, the developed FIS model development and optimization framework has some limitations. The MFs derived from the FCM clusters were approximated using Gaussian membership functions, and exploring other MFs functions such as triangular or trapezoidal is recommended. Also, during the membership function optimization process, the initial solutions and the overlaps among MFs of the final best solution were evaluated using expert judgment. Thus, collecting additional data to establish solutions and further expand the number of data instances for optimization, and introducing additional constraints on the standard deviation parameter of the MFs to improve the interpretability of optimized MFs are recommended so as to improve the MFs optimization process.

8.3.3: CLP Models Application

Construction labour productivity models allow us to imitate part of the real construction world and provided an understanding of the current situation of CLP using the influencing factors, practices, and work sampling proportions. In this study, the developed CLP models were mainly used for predicting productivity. However, CLP models can also be applied to test the potential effects of future options and facilitate the optimization of CLP. CLP models can provide valuable information for evaluating scenarios and for development of productivity improvement strategies.

Accordingly, further application of the developed models is needed in order to achieve the following tasks: (1) apply and evaluate the accuracy of the developed models in estimating labour productivity values for upcoming projects, and further improve the developed model's performance using field CLP values, (2) develop and analyze construction project scenarios, based on relevant combinations of the CLP influencing parameters (e.g. crew size, craftsperson training level, season of construction, superintendent experience level, direct work proportion, etc.) and test the usefulness of the developed models in evaluating such scenarios, and (3) carry out optimization analysis using developed models and the associated individual and combinations of the identified key influencing parameters (factors, practices, and work sampling proportions) and evolutionary or population based optimization techniques, and propose best practices that can lead to improved CLP. Such best practices will be based on the optimal values of the various influencing parameters under various contexts, established using advanced optimization analysis with the help of the developed context-specific CLP models.

8.4: REFERENCES

Busygin, S., Oleg P., and Panos M. P. (2008). "Biclustering in data mining." *Comput. Oper. Res.* 35(9), 2964-2987.

Madeira, S. C., and Oliveira, A. L. (2004). “Biclustering algorithms for biological data analysis: a survey.” *IEEE/ACM T. Comput. Biol. Bioinformat.*,1(1), 24-45.

Pinto, A., Nunes, I. L., and Ribeiro, R. A. (2011). “Occupational risk assessment in construction industry–Overview and reflection.” *Safety Sci.*, 49(5), 616-624.

Bibliography

Aamodt, A., and Plaza, E. (1994). "Case-based reasoning: foundational issues, methodological variations, and system approaches." *AI Commun.,* 7, 39–59.

AbouRizk, S., Knowles, P., and Hermann, R. (2001). “Estimating labor production rates for industrial construction activities.” *J. Constr. Eng. Manage*., 127(6), 502–551.

Aft, L.S. (2000). “Work sampling.” *Work measurement and methods improvement*, 1st Ed., John Wiley and Sons, New York, US, 299–324.

Ahmad, S. S. S., Ahmad, S., and Pedrycz, W. (2012). “Data and feature reduction in fuzzy modeling through particle swarm optimization.” *Appl. Comput. Intell. Soft Comput.,* 2012, 21 pp.

Alberta Apprentice (2012). “Apprentice training.” <http://www.tradesecrets.gov.ab.ca/index.html?page=setting_industry_standards/training_providers.html> (May 04, 2012).

Al–Ghamdi, M.A. (1995). “Evaluation of work sampling as an indicator of construction labour productivity.” M.Sc. thesis, King Fahd Univ. of Petroleum and Minerals, Dhahran, Saudi Arabia.

Anand, K. B., and Ramamurthy, K. (2003). "Laboratory-based productivity study on alternative masonry systems." *J. Constr. Eng. Manage*., 129(3), 237–242.

Angelov, P., and Buswell, R. (2002). “Identification of evolving fuzzy rule-based models.” *IEEE T. Fuzzy Syst.*, 10(5), 667–677.

Ankrah, N.A. (2007). “An investigation into the impact of culture on construction project performance.” PhD thesis, Univ. of Wolverhampton, UK.

Archiszewski, T., Skolicki, Z., and De Jong, K., 2005. “Intelligent agents: fundamentals.” *Agents and multi-agent systems in construction*, Anumba, C.J., Ugwu, O.O., and Ren, Z., New York, US, 6–30.

Awad, A. (2012). “Contractor prequalification using hybrid systems.” PhD thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Awad, A., and Fayek, A. Robinson. (2012). "A decision support system for contractor prequalification for surety bonding." *Autom. Constr.,* 21, 89–98.

Awad, A., and Fayek, A. R. (2013). "Adaptive learning of contractor default prediction model for surety bonding." *J. Constr. Eng. Manage*., 139(6), 694–704.

Bandiera, O., Barankay, I., and Rasul,I. (2004). “Cooperation in the workplace: Evidence from the field.” *London Sch. Econo., Working paper*.

Bazire, M., and Brézillon, P. (2005). “Understanding context before using it.” *Modeling and using context,* Springer, New York, US, 29–40.

Botta, A. (2008). “Automatic context adaptation of fuzzy systems.” Ph.D. thesis, IMT Institute for Advanced Studies, Lucca, Italy.

Botta, A., Lazzerini, B., Marcelloni, F., and Stefanescu, D. C. (2009). "Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index." *Soft Compu.*, 13(5), 437–449.

Brézillon, P., and Pomerol, J. (2001). “Modeling and using context for system development: Lessons from experience.” *J. Decis. Systems,* 10(2), 265–288.

Busygin, S., Oleg P., and Panos M. P. (2008). "Biclustering in data mining." *Comput. Oper. Res.* 35(9), 2964-2987.

Campbell, D.J. (1988). “Task complexity: A review and analysis.” *Acad. Manage. Rev.*, 13 (1), 40–52.

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, A., Scott, D., and Chan, P. (2004). “Factors affecting the success of a construction project.” *J. Constr. Eng. Manage*., 130(1), 153–155.

Chang, A., and Ibbs, W. (2006). “System model for analyzing design productivity.” *J. Manage. Eng.*, 22(1), 27–34.

Chang, L.M. (1986). “Inferential statistics for craftsman questionnaire.” *J. Constr. Eng. Manage*., 112(4), 492–499.

Chang, L.M., and Borcherding, J.D. (1985). “Evaluation of craftsman questionnaire.” *J. Constr. Eng. Manage*., 111(4), 426–437.

Chanmeka, A., Thomas, S., Caldas, C., and Mulva, S. (2012). “Assessing key factors impacting the performance and productivity of oil and gas projects in Alberta.” *Can. J. Civ. Eng.*, 39(3), 259–270.

Christian, J., and Hachey, D. (1995). "Effects of delay times on production rates in construction." *J. Constr. Eng. Manage*., 121(1), 20–26.

CII (Construction Industry Institute). (1990). *Productivity measurement: An introduction,* Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2006). *Work force view of construction labor productivity*, Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2010). *Guide to activity analysis*, Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2011). *Productivity benchmarking summary report*, Univ. of Texas at Austin, Austin, Texas, US.

CII (Construction Industry Institute). (2013). *The construction productivity handbook*, Univ. Texas at Austin, Austin, Texas, US.

COAA (2012). “Benchmarking: Contractor Questionnaire.” *Construction Owners Association Alberta*, Version 8.4, Edmonton, Alberta, Canada.

Cordón, O. (2011). “A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems.” *Intern. J. Approx. Reasoning,* 52(6), 894–913.

CSC (Construction Sector Council). (2013). *Construction looking forward, national summary, 2013–2021 key highlights*, Construction Sector Council, Ottawa, Ontario, Canada.

Dai, J., and Goodrum, P. M. (2012). “Generational differences on craft workers' perceptions of the factors affecting labour productivity.” *Can. J. Civ. Eng.*, 39(9), 1018–1026.

Dai, J., Goodrum, P. M., and Maloney, W. F. (2007). “Analysis of craft workers' and foremen's perceptions of the factors affecting construction labour productivity.” *Constr. Manage. Econ.,* 25(11): 1139–1152.

Dai, J., Goodrum, P. M., and Maloney, W. F. (2009). “Construction craft workers’ perceptions of the factors affecting their productivity.” *J. Constr. Eng. Manage*., 135(3), 217–226.

Dai, J., Goodrum, P., Maloney, W., and Srinivasan, C. (2009). "Latent structures of the factors affecting construction labor productivity." *J. Constr. Eng. Manage*., 135(5), 397–406.

Da Silva, S. P., and Ruwanpura, J. (2006). "Concreting productivity improvements in Alberta on commercial construction projects." *Proc., International Specialty Conference,* CSCE, Calgary, Alberta, Canada, CT-029, 1–8.

Dick, S., and Kandel, A. (2001). "Granular computing in neural networks." *Granular Computing: An Emerging Paradigm*, New York: Physica-Verlag Heidelberg, US, 275–305.

Dissanayake, M., Fayek, A. Robinson, Russell, A. D., and Pedrycz, W. (2005). “A hybrid neural network for predicting construction labour productivity.” *Proc., Intern. Computing in Civil Eng.,* ASCE*,* Cancun, Mexico, 1–12.

Dozzi, S.P., and AbouRizk, S.M. (1993). “Productivity in construction.” *Institute for Research in Construction*, National Research Council, Ottawa, Ontario, Canada.

Drewin, F.J. (1982). *Construction productivity: measurement and improvement through work study,*1st Ed., Elsevier Science Publishing Inc.

Durdyev, S., and Mbachu, J. (2011). "On-site labour productivity of New Zealand construction industry: Key constraints and improvement measures." *Aust. J. Constr. Econ. Building*, 11(3), 18–33.

El-Gohary, K.M., and Aziz, R.F. (2014). “Factors influencing construction labor productivity in Egypt.” *J. Manage. Eng.*, 30(1), 1–9.

El-Rayes, K., and Mosehli, O. (2001). “Impact of rainfall on the productivity of highway construction.” *J. Constr. Eng. Manage.,* 127(2), 125–131.

Engwall, M. (2003). "No project is an island: linking projects to history and context." *Res.Pol.,* 32(5), 789–808.

Eslamdoost, E., and Heravi, G. (2013). “Identification and evaluation of effective factors on labor productivity in power plant construction projects.” *Proc. Construction Specialty Conference*, CSCE, Montreal, Quebec, Canada, 10 pp.

Ezeldin, A. S., and Sharara, L. M. (2006). "Neural networks for estimating the productivity of concreting activities." *J. Constr. Eng. Manage*., 132(6), 650–656.

Fayek, A. Robinson. (2011). *Fuzzy hybrid modeling for strategic construction analysis and delivery*. Natural Sciences and Engineering Research Council of Canada Industrial Research Chair proposal, Dept. of Civil and Environmental Engineering, Univ. of Alberta, Edmonton, Alberta, Canada.

Fayek, A. Robinson, Dissanayake, M., and Campero, O. (2003). “Measuring and classifying construction field rework: A pilot study.” *Construction Owners Association Alberta*, Field Rework Committee, Edmonton, Alberta, Canada.

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

Fayek,A. Robinson, and Poveda, C. (2008). “A pilot study to develop a skills development tool for construction trades foremen.” *COAA*, Supervisory Training and Qualifications, Edmonton, Alberta, Canada.

Fayek, A. Robinson, Shaheen, A., and Oduba, A. (2002). *A pilot study to examine the effective integration of apprentices into the industrial construction sector*, Dept. of Civil and Environmental Engineering, Univ. of Alberta. Presented to the Workforce Development Coordinating Committee, Construction Owners Association of Alberta, Edmonton, Alberta, Canada.

Fayek, A. Robinson, and Tsehayae, A. A. (2012). "Modeling construction labour productivity using fuzzy logic and exploring the use of fuzzy hybrid techniques." *Fuzzy Information Processing Society (NAFIPS), 2012 Annual Meeting of the North American,* IEEE, Berkeley, California, US, 1–6.

Frontline (2015). "Introduction: Genetic algorithms and evolutionary algorithms." <http://www.solver.com/genetic-algorithm> (March 16, 2015).

Fulford, R., and Standing, C. (2014). "Construction industry productivity and the potential for collaborative practice." *Int. J. Project Manage.*, 32(2), 315–326.

Gacto, M. J., Alcalá, R., and Herrera, F. (2011). "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures." *Inf. Sci.,* 181(20), 4340–4360.

Gerek, İ. H., Erdis, E., Mistikoglu, G., and Usmen, M. (2014). "Modelling masonry crew productivity using two artificial neural network techniques." *J. Civ. Eng. Manage.*, (ahead-of-print), 8pp.

Gouett, M. C., Haas, C. T., Goodrum, P. M., and Caldas, C. H. (2011). "Activity analysis for direct-work rate improvement in construction." *J. Constr. Eng. Manage*., *137*(12), 1117–1124.

Grau, D., Caldas, C. H., Haas, C.T., Goodrum,P. M., and Gong, J. (2009). "Assessing the impact of materials tracking technologies on construction craft productivity." *Autom. Constr.,* 18(7): 903–911.

Gray, A.R., and MacDonell, S.G. (1997). "A comparison of techniques for developing predictive models of software metrics." *J. Inf. and Software Tec.*, 6(39), 425–437.

Green, S. (1991). "How many subjects does it take to do a regression analysis?" *Multivar. Behav. Res.*, 26, 499–510.

Green, S. D., Kao, C., and Larsen, G. D. (2009). "Contextualist research: Iterating between methods while following an empirically grounded approach." *J. Constr. Eng. Manage.,* 136(1), 117–126.

Gudwin, R.R., and Gomide, F. (1994). "Context adaptation in fuzzy processing." *Proc., Brazil-Japan Joint Symposium on Fuzzy Systems*, Campinas, Brazil.

Gudwin, R. R., Gomide, F. A., and Pedrycz, W. (1998). "Context adaptation in fuzzy processing and genetic algorithm." *Int. J. Intell. Syst*., 13(10-11), 929–948.

Guyon, I. and Elisseeff, A. (2003). “An introduction of variable and feature selection.” *J. Mach. Learn. Res., 3*, 1157–1182.

Hall, M. A. (1998). “Correlation-based feature selection for machine learning.” PhD thesis, Univ. of Waikato, Hamilton, New Zealand.

Hallowell, M. R. (2010). “Worker fatigue: Managing concerns in rapid renewal highway construction projects.” *Prof. Safety*, 55(12), 18–26.

Halpin, D. W. (2007). “Fifty years of progress in construction engineering research.” *J. Constr. Eng. Manage.,* 133(9), 635-639.

Handa, V.K., and Abdalla, O. (1989). “Forecasting productivity by work sampling.” *J. Cons. Mgmt. Econ*., 7, 19–28.

Hanna, A. S., (2010). *Construction labour productivity management and methods improvement*. n.p., Madison, Wisconsin, US.

Hanna, A. S., Chang, C.K., Sullivan, K.T., and Lackney, J.A. (2008). "Impact of shift work on labor productivity for labor intensive contractor." *J. Constr. Eng. Manage*.,134(3), 197–204.

Hanna, A. S., Russel, J., Gotzion, T., and Nordheim, E. (1999a). “Impact of change orders on labor efficiency for mechanical construction.” *J.Constr. Eng. Manage.*, 125(3), 176–184.

Hanna, A. S., Russel, J., Nordheim, E., and Bruggink, M. (1999b). “Impact of change orders on labor efficiency for electrical construction.” *J. Constr. Eng. Manage.*, 125(4), 224–232.

Hanna, A. S., Taylor, C. S., and Sullivan, K. T. (2005). "Impact of extended overtime on construction labor productivity." *J. Constr. Eng. Manage*., 131(6), 734–739.

Harrison, P. (2007). “Can measurement error explain the weakness of productivity growth in the Canadian construction industry?” *Centre for the Study of Living Standards*, Ottawa, Ontario, Canada.

Hart, H. (1995).” Measuring construction productivity.” *Participant handbook, Construction Industry Institute*, Austin, Texas, US.

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press, New York, US.

Heller, F. A., and Yukl, G. (1969). “Participation, managerial decision-making, and situational variables.” *Organ. Behav. Human Perfor.*, 4(3), 227–241.

Hendrickson, C., Martinelli, D., and Rehak,D. (1987). "Hierarchical rule-based activity duration estimation," *J. Constr. Eng. Manage*., 113 (2), 288–301.

Heravi, G., and Eslamdoost, E. (2015). "Applying artificial neural networks for measuring and predicting construction-labor productivity." *J. Constr. Eng. Manage*.,04015032.

Herbsman, Z., and Ellis, R. (1990). “Research of factors influencing construction productivity.” *Constr. Manage. Econ*., 8(1), 49–61.

Ho, D. T. (2013). “Context dependent fuzzy modelling and its application.” PhD thesis, Univ. of Nottingham, Nottingham, UK.

Horman, M. J., and Kenley, R. (2005). "Quantifying levels of wasted time in construction with meta-analysis." *J. Constr. Eng. Manage*., 131 (1), 52–61.

HSE (2005). “Noise at work: Guidance for employers on the control of noise at work.” Health and Safety Executive, INDG362, Version 1.

HSE (2012). Health and Safety Executive, <http://www.hse.gov.uk/pubns/iacl95.htm> (May 05, 2012).

Huh, Y. 2004. “Highway bridge construction production rates for time estimation.” PhD thesis, Univ. of Texas at Austin, Austin, Texas, US.

Ibbs, C. (1997). “Quantitative impacts of project change: Size issues.” *J. Constr. Eng. Manage.*, 123(3), 308–311.

IMF (2015). “World economic outlook.” International Monetary Fund, < http://www.imf.org/external/pubs/ft/weo/2015/01/ > (June 28, 2015).

Jang, H., Kim, K., Kim, J., and Kim, J. (2011). “Labour productivity model for reinforced concrete construction projects.” *J. Cons. Innov.: Info. Proc. Manage.*, 11(1), 92–113.

Jang, S., and Woo, W. (2003). “Ubi-UCAM: A unified context-aware application model.” *Modeling and Using Context,* Springer, New York, US, 178–189.

Jarkas, A., and Bitar, C. (2012). “Factors affecting construction labor productivity in Kuwait.” *J. Constr. Eng. Manage*., 138(7), 811–820.

Jarkas, A., and Horner, M. (2011). "Revisiting the applicability of learning curve theory to formwork labour productivity." *Constr. Manage. Econ.,* 29(5), 483–493.

Jayawardane, A. K. W., Price, A. D., and Harris, F. C. (1995). “Measurement of brickwork and blockwork productivity: Part A.” *Building Res. Inform.*, 23(2), 81–86.

Jergeas, G. (2009). “Improving construction productivity on Alberta oil and gas capital projects.” Alberta Finance and Enterprise, Alberta, Canada.

Ji, S., Park, M., and Lee, H. (2012). "Case adaptation method of case-based reasoning for construction cost estimation in Korea." *J. Constr. Eng. Manage*., 138(1), 43–52.

Josephson, P., and Björkman, L. (2013). “Why do work sampling studies in construction? The case of plumbing work in Scandinavia.” *Eng. Construct. Architect. Manage.*, 20(6), 589–603.

Keith, T. Z. (2015). Multiple regression and beyond: An introduction to multiple regression and structural equation modeling, 2nd Ed., Routledge, New York, US.

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

Koehn, E., and Brown, G. (1985). “Climatic effects on construction.” *J. Constr. Eng. Manage.*, 111(2),129–137.

Knight, K., and Fayek, A. Robinson (2000). “A preliminary study of the factors affecting the cost escalation of construction projects.” *Can. J. Civ. Eng.*, 27(1), 73–83.

Knowles, R. P. (1997). “Predicting construction labour productivity using neural networks.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Leite, D., Ballini, R., Costa, P., and Gomide, F. (2012). “Evolving fuzzy granular modeling from nonstationary fuzzy data streams.” *Evolving Sys.*, *3*(2), 65–79.

Liberda, M., Ruwanpura, J., and Jergeas, G. (2003). "Construction productivity improvement: A study of human, management and external issues." *Proc., Construction Research Congress*, ASCE, Honolulu, Hawaii, 1–8.

Lim, E. C., and Alum, J. (1995). "Construction productivity: issues encountered by contractors in Singapore." *Inter. J. Proj. Manage*., 13(1), 51–58.

Lin, C., and Huang, H., 2010. "Improved baseline productivity analysis technique." *J. Constr. Eng. Manage*., 136(3), 367–376.

Liou, F., and Borcherding, J. D. (1986). "Work sampling can predict unit rate productivity." *J. Constr. Eng. Manage*., 112(1), 90–103.

Liu, H., and Yu, L. (2005). "Toward integrating feature selection algorithms for classification and clustering." *IEEE T. Knowl. Data Eng.,* 17(4), 491–502.

Lu, M. (2001). "Productivity studies using advanced ANN models." PhD thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Lu, M., AbouRizk, S., and Hermann, U. H. (2000). "Estimating labor productivity using probability inference neural network." *J. Comput. Civ. Eng.,* 14(4), 241–248.

Madeira, S. C., and Oliveira, A. L. (2004). "Biclustering algorithms for biological data analysis: a survey." *IEEE/ACM T. Comput. Biol. Bioinformat.*,1(1), 24-45.

Magdalena, L. (1997). "Adapting the gain of an FLC with genetic algorithms." *Int. J. Approx. Reasoning,* 17(4), 327–349.

Magdalena, L. (2002). "On the role of context in hierarchical fuzzy controllers." *Int. J. Intell. Syst.,* 17(5), 471–493.

Magdalena, L. (2015). "Fuzzy rule-based systems." *Springer Handbook of Computational Intelligence*. Springer Berlin Heidelberg, 203–218.

Mao, H. (1999). “Estimating labour productivity using fuzzy set theory.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Mencar, C., Castellano, G., and Fanelli A.M. (2007). "Distinguishability quantification of fuzzy sets." *Inform. Sci.*, 177, 130–149.

Molleman, E., and van den Beukel, A. (2007). “Worker flexibility and its perceived contribution to performance: The moderating role of task characteristics.” *Hum. Factors Ergonom. in Manufac. Service Industr.*, 17(2), 117–135.

Moosavi, S.F., and Moselhi,O. (2012). “Schedule assessment and evaluation.” *Proc., Construction Research Congress*, ASCE, West Lafayett, Indiana, US, 535–544.

Moselhi, O., Aly, R., and Hassanein, A. (2011). “Labour productivity in building construction: a field study.” *Proc., 3rd International / 9th Construction Specialty Conference*, Ottawa, Ontario, Canada.

Mosehli, O., Assem, I., and El-Rayes, K. (2005). Change orders impact on labor productivity.” *J. Constr. Eng. Manage.,* 131(3), 354–359.

Moselhi, O., Charles, L., and Fazio, P. (1991). “Impact of change orders on construction productivity.” *Can. J. Civ. Eng.*, 18(4), 484–492.

Moselhi, O., and Khan, Z. (2012). “Significance ranking of parameters impacting construction labor productivity.” *J. Cons. Innov.: Info., Proc., Manage.*, 12(3), 272–296.

Mourgues, C., and Fischer, M. (2008). “A work instruction template for cast-in-place concrete construction laborers.” *CIFE*, Working paper, 109, Stanford Univ., California, US.

Nasir, H., Ahmed, H., Haas, C., and Goodrum, P.M. (2014). “An analysis of construction productivity differences between Canada and the United States.” *Constr. Manage. Econ.,* 32 (6), 595–607.

Nasirzadeh, F., and Nojedehi, P. (2013). "Dynamic modeling of labor productivity in construction projects." *Int. J. Project Manage.,* 31(6), 903–911.

Ng, S. T., Skitmore, R. M., Lam, K. C., and Poon, A. W. (2004). “Demotivating factors influencing the productivity of civil engineering projects.” *Inter. J. Proj. Manage.*, *22*(2), 139–146.

Oduba, A. O. (2002). “Predicting industrial construction productivity using fuzzy expert systems.” M.Sc. thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

O’Connor, J. T., and Huh, Y. (2006). “Crew production rates for contract time estimation: Beam erection, deck and rail of highway bridges.” *J. Constr. Eng. Manage*., 132(4), 408–415.

Oglesby, C. H., Parker, H. W., and Gregory, A. H. (1988). *Productivity improvement in construction*. McGraw-Hill, New York, New York, US.

Oral, E. L., and Oral, M. (2010). "Predicting construction crew productivity by using self-organizing maps." *Autom. Constr.,* 19(6), 791–797.

Oral, M., Oral, E. L., and Aydın, A. (2012). "Supervised vs. unsupervised learning for construction crew productivity prediction." *Autom. Constr.,* 22, 271–276.

Panas, A., and Pantouvakis, J. P. (2010). “Evaluating research methodology in construction productivity studies.” *Built Hum. Env. Rev.,* 3(1), 63–85.

Panthi, K., Farooqui, R. U., and Ahmed, S. M. (2008). “An investigation of the leadership style of construction managers in South Florida.” *J. Const. Manage. Econ.*, 11, 455–565.

Park, H.S. (2002). “Development of a construction productivity metrics system (CPMS).” Ph.D. thesis, Univ. of Texas, Austin, Texas, US.

Pedrycz, W. (2011). "From fuzzy models to granular fuzzy models." *Fuzzy Logic and Applications*: *Proc., 9th International Workshop, WILF,* Trani, Italy, 75–82.

Pedrycz, W., and Gomide, F. (2007). *Fuzzy systems engineering: Toward human-centric computing.* John Wiley and Sons, Hoboken, New Jersey, US.

Pedrycz, W., Gudwin, R., and Gomide, F. (1997). "Nonlinear context adaptation in the calibration of fuzzy sets." *Fuzzy Sets Syst*., 88(1), 91–97.

Pedrycz, W., Russo, B., and Succi, G. (2012). "Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity." *Appl. Soft Comput.,* 12(8), 1985–1995.

Pedrycz, W., and Song, M. (2012). “Granular fuzzy models: A study in knowledge management in fuzzy modeling.” *Intern. J. Approx. Reasoning*, 53(7), 1061–1079.

Picard, H.E. (2004). “Driving down construction project labour cost,” *Proceeding, Structural and Construction Engineering*, Rome, Italy.

Pinto, A., Nunes, I. L., and Ribeiro, R. A. (2011). “Occupational risk assessment in construction industry–Overview and reflection.” *Safety Sci.*, 49(5), 616-624.

PMI (Project Management Institute). (2007). *Construction extension to the PMBOK (Project management body of knowledge) guide*. Newtown Square, Pennsylvania, US.

Portas, J., and AbouRizk, S. (1997). "Neural network model for estimating construction productivity." *J. Constr. Eng. Manage*., 123(4), 399–410.

Pulkkinen, P., and Koivisto, H. (2010). "A dynamically constrained multiobjective genetic fuzzy system for regression problems." *IEEE T. Fuzzy Syst.,* 18, 161–177.

Reyes-Galaviz, O. F., and Pedrycz, W. (2015). “Granular fuzzy models: Analysis, design, and evaluation.” *Intern. J. Approx. Reasoning*, 64, 1–19.

Richard, F., and Liu, A. (2008). *Research methods for construction.* 3rd Ed., Blackwell Publishing Ltd, West Sussex, U.K.

Rivas, R. A., Borcherding, J. D., Gonzalez, V., and Alarcon, L. F. (2011). "Analysis of factors influencing productivity using craftsmen questionnaires: Case study in a Chilean construction company." *J. Constr. Eng. Manage*., 137(4), 312–320.

Rodrigues, D., Prata, D., and Silva, M. (2015). “Exploring social data to understand child labor.” *Int. J. Social Sci. Humanity*, 5, 29–33.

Saeys, Y., Inza, I., and Larranaga, P. (2007). “A review of feature selection techniques in bioinformatics.” *Bioinformatics,* 23(19), 2507–2517.

Sandbhor, S., and Botre, R. (2014). "Applying total interpretive structural modeling to study factors affecting construction labour productivity." *Aust. J. Constr. Econ. Build.,* 14(1), 20–31.

Sanders, S. R., and Thomas, H. R. (1993). "Masonry productivity forecasting model." *J. Constr. Eng. Manage*., 119(1), 163–179.

Shahtaheri, M., Nasir, H., and Haas, C. T. (2014). "Setting baseline rates for on-site work categories in the construction industry." *J. Constr. .Eng. Manage.,* 141(5), 04014097.

Sheppard, B., and Lewicki, R. J. (1987). “Toward general principles of managerial fairness.” *Soc. Just. Res.*, 1(2), 161–176.

Silva, S. P., and Ruwanpura, J. Y. (2006). “Concreting productivity improvements in Alberta on commercial construction projects.” *Proc., 1st Int. Constr. Specialty Conf.*, Canadian Society for Civil Engineering, Calgary, Alberta, Canada.

Song, L., and AbouRizk, S. M. (2008). "Measuring and modeling labor productivity using historical data." *J. Constr. Eng. Manage*., 134(10), 786–794.

Sonmez, R. (1996). “Construction labor productivity modeling with neural networks and regression analysis.” PhD thesis, Iowa State Univ., Ames, Iowa, US.

Sonmez, R., and Rowings, J. E. (1998). "Construction labor productivity modeling with neural networks." *J. Constr. Eng. Manage*., 124(6), 498–504.

Srinavin, K., and Mohamed, S. (2003). "Thermal environment and construction workers’ productivity: some evidence from Thailand." *Build.Environ.,* 38(2), 339–345.

Statistics Canada, (2015a). “Gross domestic product at basic prices, by industry, 2010-2014.” < http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/manuf10-eng.htm> (June 16, 2015).

Statistics Canada, (2015b). “Employment by industry, 2010-2014.” < http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ40-eng.htm> (June 16, 2015).

Stewart, T. J., and Frye, A. W. (2004). “Investigating the use of negatively-phrased survey items in medical education settings: Common wisdom or common mistake?”, *Acade. Medicine*, 79(10), S1–S3.

Thomas, H. R. (1991). "Labor productivity and work sampling: the bottom line." *J. Constr. Eng. Manage*., 117(3), 423–444.

Thomas, H. R., Guevara, J. M., and Gustenhoven, C. T. (1984). “Improving productivity estimates by work sampling.” *J. Constr. Eng. Manage*., 110 (2), 178–188.

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). "Modeling construction labor productivity." *J. Constr. Eng. Manage*., 116(4), 705–726.

Thomas, H. R., and Mathews, C. T. (1986). "Learning curve models of construction productivity" *J. Constr. Eng. Manage*., 112 (2), 245–258.

Thomas, H. R., and Raynar, K. A. (1997). "Scheduled overtime and labor productivity: quantitative analysis." *J. Constr. Eng. Manage*., 123(2), 181–188.

Thomas, H. R., and Sakarcan, A. S. (1994). "Forecasting labor productivity using factor model." *J. Constr. Eng. Manage*., 120(1), 228–239.

Thomas, H.R., and Yiakoumis, I. (1987). "Factor model of construction productivity." *J. Constr. Eng. Manage.,* 113(4), 623–639.

Trost, S., and Oberlender,G. (2003). "Predicting accuracy of early cost estimates using factor analysis and multivariate regression." *J. Constr. Eng. Manage.*, 129(2), 198–204.

Tsehayae, A. A., and Fayek, A. Robinson. (2012). "A research framework for work sampling and its application in developing comparative direct and support activity proportions for different trades," *Proc., Construction Research Congress*, ASCE, West Lafayette, Indiana, US, 1420–1429.

Tsehayae, A. A., and Fayek, A. Robinson. (2014). "Data-driven approaches to discovering knowledge gaps related to factors affecting construction labour productivity," *Proc., Construction Research Congress*, ASCE, Atlanta, Georgia, US, 837–891.

Tsehayae, A.A., and Fayek, A. R. (2014a). *Factor surveys: Interview questionnaires, participation letter, and consent form*. Technical report submitted to Suncor Energy, NSERC IRC SCMD: Productivity/Suncor/ Factor Surveys Report/2014–TD01, Edmonton, Alberta, Canada, October 10, 2014, 40 pp.

Tsehayae, A.A., and Fayek, A. R. (2014b). *Data collection protocol: Labour productivity analysis and modeling*. Technical report submitted to Suncor Energy, NSERC IRC SCMD: Productivity/Suncor/ Data Collection Protocol Report/2014–TD02, Edmonton, Alberta, Canada, October 14, 2014, 99 pp.

Tsehayae, A. A., and Fayek, A. Robinson. (2014c). "Identification and comparative analysis of key parameters influencing construction labour productivity in building and industrial projects." *Can. J. Civ. Eng.,* 41(10), 878–891.

Tsehayae, A. A., Pedrycz, W., and Fayek, A. Robinson (2013). "Application of granular fuzzy modeling for abstracting labour productivity knowledge bases." *Proc., IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint Congress,* IEEE, Edmonton, Alberta, Canada, 1096–1101.

Turner, R. M. (2014). “Context-mediated behavior.” *Context in computing,* Springer, New York, US, 523–539.

Turpin, M. P., and Kamath, A. R. R. (1986). “The development of an equipment availability reporting database and analysis package.” *Relia. Eng.*, 15(2), 95–113.

Walker, A. (2015). *Project management in construction.* Blackwell Publishing Ltd, Oxford, UK.

Watkins, M., Mukherjee, A., Onder, N., and Mattila, K. (2009). "Using agent-based modeling to study construction labor productivity as an emergent property of individual and crew interactions." *J. Constr. Eng. Manage*., 135(7), 657–667.

Weiss, D.H. (2004). *Fair, square and legal: safe hiring, managing and firing practices to keep you and your company out of court*. 4 ed., AMACOM, American Management Association, New York, US.

Willmott, C., Robeson, S., and Matsuura K. (2012). “A refined index of model performance.” *Int. J. Climatol.*, 32, 2088–2094.

Xing, E. P., Jordan, M. I., and Karp, R. M. (2001). “Feature selection for high-dimensional genomic microarray data.” *In ICML*, 1, 601–608.

Xu, L., Yan, P., and Chang, T. (1988). “Best first strategy for feature selection.” *Proc. International Conference on Pattern Recognition*, IEEE, 706–708.

Xue, X., Shen, Q., Wang, Y., and Lu, J. (2008). “Measuring the productivity of the construction industry in China by using DEA-based Malmquist productivity indices.” *J. Constr. Eng. Manage*. 134(1), 64–71.

Yao, J. (2007). "A ten-year review of granular computing." *Proc., Intl. Conf. on Granular Computing,* IEEE, Silicon Valley, California, US, 734–739.

Yi, W., and Chan, A. (2014). "Critical Review of Labor Productivity Research in Construction Journals." *J. Manage. Eng.,* 30(2), 214–225.

Zadeh, L.A. (1998). "Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems." *Soft Comput.*, 2, 23–25.

Zadeh, L.A. (1997). "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic." *Fuzzy Sets and Systems,* 90, 111–117.

Zadeh, L. A. (2008). "Is there a need for fuzzy logic?" *Inf. Sci.,* 178(13), 2751–2779.

Zemouri, R., Gouriveau, R., and Zerhouni, N. (2010). “Defining and applying prediction performance metrics on a recurrent NARX time series model.” *Neurocomputing*, 73(13–15), 2506–2521.

Zhu, Y. (2005). "Applying context to systems integration in AEC." *Proc., Constr. Research Congress*, ASCE, San Diego, California, US, 1–8.

Appendices

Appendix A: Input Parameters (Factors and Practices) Influencing CLP

A.1: Summary of Past Studies Identifying CLP Influencing Factors

Study Details	Categories	Identified factors
Lim and Alum (1995) Country: Singapore 17 Total Factors	Manpower	Recruitment of supervisors; Recruitment of workers; Labour turnover; Absenteeism; Communications problems with foreign workers; Alcoholism and similar problems among workforce; Labour disruptions
	Management	Materials shortages; Delays in materials deliveries to site; Disruption of power/water supplies; Stop-work orders because of site accidents; Stoppages because of work being rejected by consultants; Stop-work orders because of infringements of government regulations; Stoppages because of disputes with owners/consultants; Stoppages because of insolvency of subcontractors/suppliers
	Environment	Health; Inclement weather
Oduba (2002) Country: Canada 57 Total Factors	Factors	Project-level [Extent of fast tracking; Criticality of schedule; Tightness of budget], Activity-level attribute [Pipe length; Pipe diameter; Complexity of shape of pipe; Efficiency of rigging method; Wall thickness or schedule; Shelter requirement; Purge requirement; Pre-heat requirement; Bevel dimension or joint configuration; Crew ratio; Task crew size; Overall crew size; Elevation; Impact of weather conditions; Ground conditions; Access to work area; Crowding of work area; Adequacy of site storage; Sufficiency of number of crew members; Crew skill level; Crew turnover; Average temperature; Average wind speed; Average precipitation; Crew experience (learning); Crew experience (seniority); Number of consecutive days worked; Scaffold requirement; Average relative humidity; Amount of rework; Amount of change orders; Drawings and specifications quality; Extent and quality of training; Extent and quality of supervision; Number of disruptions per day; Percentage overtime per week; Frequency and extent of material shortages; Magnitude of organizational constraints; Inspection requirements; Safety requirements; Quality requirements; Percentage of prefabricated or modularized work; Equipment availability]
	Context Variables	Project-level [Project location; Province; Year of construction; Client; Contract type; Project definition; Project type; Union status; Project sector; Season of construction; Location of work scope], Activity-level [Material type; Weld type; Filler material type]
Liberda et al. (2003) Country: Canada 51 Total Factors	Human	Worker motivation; Worker boredom and fatigue; Worker attitude and morale; Workers physical limitations; Worker absenteeism; Team-spirit of the crew; Worker learning curve; Worker experience and skills
	External	Union rules and influences; Adverse weather conditions; Noise; dust; radiation; Congested work area; Changes in drawings and specifications; Changes in contract; Demand of over-quality work; Nature of project (size and complexity)
	Management	Protective gear; Unrealistic schedule; Overtime; Multiple shifts; Excessive shift length; Disrespectful treatment of worker; Parking facilities; Salary and benefits; Site layout; Necessity to re-do work; Discontinuity in crew makeup; Failure to utilise workers skills; Incompetent personnel; Overcrowded work areas; Poor inspection programs; Unsafe working conditions; Inadequate equipment; Inadequate supervision; Composition of the crew; Constructability; Out-of-sequence work; Interruption and disruption; Adequate site facilities for worker; Lack of co-operation between crafts; Inadequate communication; Lack of worker training and education; Cleanliness of construction site; Lack of procedures for construction methods; Subcontracting; Changes in foremen; Non availability of materials; Non availability of tools; Non availability of information; Non availability of equipment; Lack of detailed planning

Study Details	Categories	Identified factors (continued)
CII (2006) Country: US 83 Total Factors	Supervisor direction	Inadequate instruction provided; Not receiving directions due to size of the project; Receiving compliments for doing a good job; Being notified of mistakes when they occur; Lack of goals for craft workers
	Communication	Different languages spoken on a project; Disregard of crafts' productivity improvement suggestion; Lack of "Big Picture" view on behalf of the crafts; Craft worker importance; Lack of communication among site management
	Safety	Shortage of personal protective equipment; Lack of site safety resources
	Tools and consumables	Availability of consumables; Restrictive project policy on consumables; Availability of hand tools; Availability of power tools; Lack of power source for tools; Lack of extension cords; Inexperienced tool room attendants; Misplaced tools; Poor quality power tools
	Materials	Availability of material; Poor material quality; Availability of bulk commodities; Errors in prefabricated material; Difficulty in tracking material
	Engineering drawing management	Drawing errors; Availability of drawings; Slow response to questions with drawings; Drawing legibility; Needed information not on drawings
	Labour	Availability of skill training; Jobsite orientation program; Availability of health and safety training; Qualified craftsmen; Craftsmen's pride in their work; Craftsmen's incentive; Motivated craft workers; Equal pay on projects in a geographic area; Craft workers' trust in supervisors
	Foreman	Foremen people skill; Qualified foremen; Fair/just performance reviews; Foremen allowing crafts to work autonomously; Lack of construction knowledge on behalf of foremen; Lack of authority to discipline craft workers; Lack of proper resource allocation
	Superintendent	Proper managerial and administrative support; Excessive paperwork; Superintendent's people skill; Qualified superintendents; Lack of experience on behalf of superintendents; Respect for craft workers; Micromanagement on behalf of superintendent; Political/performance competitions within company; Inconsistent safety policies established by different superintendents; Different work rules by superintendents
	Project management	Delay in work permits; Out of sequence work assignments; Absenteeism; Reasonable project goals and milestones; Respect for craft workers and foremen; Layoff of qualified craft workers; Awareness of on-site activities and project progress; Pulling people off a task before it is done; Jobsite congestion; Different pay scales for the same job on a project; Different per diem rate; Incentive for good performance; Material storage area too far from workface; Insufficient size of material storage area; Shortage of temporary facilities; Coordination between the trades; Slow decisions; Correct crew size; Vehicle traffic routes; Weather protection
	Construction equipment	Availability of crane or forklift; Availability of man-lift; Waiting for people and/or equipment to move material; Poor equipment maintenance; Equipment repairs; Maintenance of power tools

Study Details	Categories	Identified factors (continued)
Durdyev and Mbachu (2011) Country: New Zealand 56 Total Factors	Internal constraints	Project Finance [Late payments; Reworks; Undervalued work/poor estimation; Dispute and litigation costs; Lenders" high interest charges; High insurance premiums; Inadequate supply or high cost of needed resources: money; labour]; Workforce [Level of commitment; Level of empowerment; Level of skill and experience; Level of familiarity with current job and conditions; Level of involvement of direct labour or subcontract; Workforce Absenteeism; Level of staff turnover/churn rate; Health of the workforce], Technology/process [Suitability or adequacy of plant and equipment; Method of construction; Technology employed; Lack of awareness of or training on new technologies; Resistance to accept new technologies; Inadequate IT infrastructure and application in construction industry], Project characteristics [Site conditions: access, subsoil, topography; Project complexity; Buildability issues; Site location and environment; Type of procurement adopted], Project management [Adequacy of planning and risk management process; Coordination, supervision, performance monitoring, and control; Project organisational culture; Relationship management; Competencies of the project team; Project management style; Frequency of design changes; Client's over influence on the construction process]
	External constraints	Statutory compliance [Health and safety in employment act; Resource management act; Local authority bylaws; Construction contracts act; Building act/building consent/building regulations; Employment relations act; Consumer guarantees act; Fair trading act; Arbitration act], Unforeseen events [Inclement weather; Ground conditions necessitating revisions; On-site accidents/ Acts of God; Natural disasters], Other external forces [Inflation/ fluctuations in material prices; Fluctuations in exchange rate; Energy crises/costs; Interest rate/cost of capital; Market conditions and level of competitions in the industry for jobs; Frequent changes in government policies/legislations impacting on construction; Rapid technological advances; Increase in industry or society-wide litigations/adversarial relations]
Jarkas and Bitar (2012) Country: Kuwait 45 Total Factors	Technological	Clarity of technical specifications; The extent of variation/change orders during execution; Coordination level among design disciplines; Design complexity level; Stringent inspection by the Engineer; Delay in responding to requests for information (RFI); Compatibility and consistency among contract documents; Rework; Site restricted access; Confinement of working space; Site layout; Inspection delay by the engineer
	Human/labour	Motivation of labour; Skill of labour; Physical fatigue; Shortage of experienced labour
	Management	Lack of labour supervision; Proportion of work subcontracted; Lack of incentive scheme; Construction manager's lack of leadership; Unsuitability of storage location; Working overtime; Crew size and composition; Unrealistic scheduling and expectation of labour performance; Labour interference and congestion; Shortage of materials; Construction method; Payment delay ;Communication problems between site management and labour; Accidents as a result of poor site safety program; Late arrival, early quit, and frequent unscheduled breaks; Unavailability of suitable tools ;Lack of training offered to operatives; Inspection delay by site management; Sequencing problems; Lack of recognition program; Lack of periodical meetings with crew leaders; Owner's representative intervention with site management and operatives; Lack of suitable rest area offered to labour on site; Lack of providing labour with transportation
	External	High/low temperature; High humidity; Sandstorms; High winds; Rain

A.2: Quantification of Activity Level Input Parameters

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.1.1	Crew Properties	Refers to the nature and property of the crew and its members which will directly involve with execution of the tasks.		D	FM
1.1.1.1	Crew size	The total size of the crew performing the actual task will have a direct effect on the amount of output.	Integer (Total number of crew members)	D	DC
1.1.1.2	Craftsperson education	Refers to the highest achieved education level of craftsperson in a crew. The most common education level of the crew members is recorded.	Categorical (Below Secondary, Secondary School, Technical or Apprentice, College, University)	C	DC
1.1.1.3	Craftsperson on job training	Craftspeople are expected to get job specific trainings to improve their skillset. Any training, for erecting scaffolding, rigging and hoisting, zoom boom operation, etc. provided to craftspeople during his/her career is recorded. The average training hour per crew is recorded.	Real number (No. trainings attended x Duration of Training, hrs.)	C	FM
1.1.1.4	Craftsperson technical training	Craftspeople are technical trained to ensure they have the necessary technical skills to perform the task. Any technical training, including apprentice trainings to fulfill the trade qualification requirements is recorded.[9] The average training hour per crew shall be recorded.	Real number (No. trainings attended x Duration of Training, hrs.)	C	FM
1.1.1.5	Crew composition	Refers to a crew composition in terms of ratio of journeyman to apprentice, which directly influences the overall experience of the crew.	Proportion (Ratio Journeyman to Apprentice)	D	FM
1.1.1.6	Crew experience (seniority)	Refers to the average years of experience of the crew members on the trade under study where higher values are expected to have improved productivity and learning speed.	Real number (Crew average years of experience for the trade under study)	C	FM
1.1.1.7	Number of languages spoken	The number of languages used on principal work communication will influence the homogeneity of the crew.	Integer (Number of languages used in site work related communication, total for a crew)	C	FM

[9] Alberta Apprentice (2012). <http://www.tradesecrets.gov.ab.ca/index.html?page=setting_industry_standards/training_providers.html> (May 04, 2012).

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.1.1.8	Co-operation among craftsperson	Refers to an overall measure of the cooperation among crew members. The measure in this study will be based on the diversity of crew members in terms of ability, stake value (the expected advantage or gain from the successful completion of the task at hand) and the size of the crew.[10] Crew members are expected to have similar nationality based on citizenship, permanent residency, or work permit.	1 - 5 Predetermined rating (shown below)	D	FM
		1. VERY DIVERSE Ability, VERY LOW Stake Value, VERY LARGE Crew Size; 2. DIVERSE Ability, LOW Stake Value, LARGE Crew size; 3. DIVERSE Ability, MEDIUM Stake Value, AVERAGE Crew Size; 4. SIMILAR Ability, HIGH Stake Value, SMALL Crew Size; 5. VERY SIMILAR Ability, VERY HIGH Stake Value, VERY SMALL Crew Size			
1.1.2	Craftsperson learning effect	Refers to the overall gain in productivity due to the effect working on repetitive tasks over a long period of time. Represented in terms of $Y = ax^b$, Y– Installation time for next unit, a - time for first unit, and b - learning curve coefficient		I	SI
1.1.2.1	Time to install the first unit (a)	Refers to the time required to install first unit of the activity on the project under study.	Real number (Time to install first unit, min)	I	SI
1.1.2.2	Learning coefficient (b)	Refers to the average time saving in percent between first and consecutive units.	Real number (%, Average time saving between first and consecutive units)	I	SI
1.1.3	Treatment of craftsperson by foreman	Self-explanatory. The measure will be based on the respectfulness (having a respect to crew members), sincerity (honest without pretending to crew members), and counseling (giving advice and support to crew members) nature of foreman.	1 - 5 Predetermined rating (shown below)	W	CM
		1. ALWAYS Disrespectful, Insincere, NO Counseling; 2. OFTEN Disrespectful, Insincere, NO Counseling; 3. SOMETIMES Respectful, SOMETIMES Sincere, SOMETIMES Counseling; 4. OFTEN Respectful, OFTEN Sincere, OFTEN Counseling; 5. ALWAYS Respectful, ALWAYS Sincere, ALWAYS Counseling			
1.1.4	Craftsperson motivation	Refers to the measurement of the craftsperson motivation using the Vroom's Expectancy theory: Motivation = E x I x V[2]		W	CM
1.1.4.1	Expectancy	Subjective probability of achieving a performance with a given level of effort.	Percentage (Subjective probability)	W	CM
1.1.4.2	Instrumentality	Subjective probability of achieving some outcomes based on a successful performance.	Percentage (Subjective probability)	W	CM
1.1.4.3	Valence	The effect of the outcomes as being positive (bonus payments) or negative (disciplinary actions).	Integer (-1 or 1)	W	CM

[10] Bandiera, O., Barankay, I., and Rasul,I. (2004). "Cooperation in the Workplace: Evidence From the Field." *London Sch. Econo., Working paper*.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.1.5	Craftsperson fatigue	Long work hours (overtime), extended work periods, and repetitive work tasks are used as leading indicators of physical fatigue[11]		W	SI
1.1.5.1.	Number of consecutive days worked per week	Consecutive days worked is used to measure the level of extended work period.	Integer (average consecutive days worked for crew per week)	W	SI
1.1.5.2.	Level of regular work time	Level of regular time is used to measure short-term physical fatigue.	Integer (Total regular work hour per week, hrs.)	W	SI
1.1.5.3.	Level of overtime	Level of overtime is used to measure long-term physical fatigue.	Integer (Total over time per week, hrs.)	W	SI
1.1.6	Craftsperson trust in foreman	Crew member's trust in the judgment and overall capacity of the foreman will lead to better performance.	1 - 5 Predetermined rating (shown below)	W	CM
		1. VERY LOW Trust; 2. LOW Trust; 3. AVERAGE Trust; 4. HIGH Trust; 5. VERY HIGH Trust			
1.1.7	Team spirit of crew	Crew member's team spirit will directly influence their performance.	1 - 5 Predetermined rating (shown below)	W	FM
		1. VERY LOW Team Spirit; 2. LOW Team Spirit; 3. AVERAGE Team Spirit; 4. HIGH Team Spirit; 5. VERY HIGH Team Spirit			
1.1.8	Level of absenteeism	Absenteeism is known to affect crew makeup, morale of workers, and labour productivity. The weekly average daily absenteeism per crew is recorded.	Percentage (daily number of absent crew members to total crew size, daily values averaged weekly)	W	FM
1.1.9	Crew turnover	Crew turnover is also known to affect crew makeup and labour productivity. The turnover of a crew member in terms of ratio of number of workers getting out of work to average weekly crew size per week is recorded.	Percentage turnover rate (% of crew turnover to average crew size per week)	W	FM
1.1.10	Discontinuity in crew makeup	Refers to the change in the makeup of a crew. Keeping a crew and its member's uniform might provide better communication and increased productivity. The change in crew makeup per day is recorded.	Percentage (% occurrence of crew member change to total crew size per day per crew)	D	FM
1.1.11	Level of interruption and disruption	To document the number of delay events caused due to several reasons, which may disrupt the crew from performing the assigned tasks.	Integer (Number of interruption and disruption events per day)	D	FM

[11] Hallowell, M. R. (2010). "Worker fatigue: Managing concerns in rapid renewal highway construction projects." *Prof. Safety*, 55(12), 18-26.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.1.12	Fairness of work assignment	Refers to the feeling of the crew members towards the assignment of work by foreman to the different crews and crewmembers. It will be measured in terms of consistency (same policy), reasonableness (use of common sense), and information (provision of information)[12]	1 - 5 Predetermined rating (shown below)	D	CM
		1. Inconsistent work assignment on a daily basis, Unreasonable work assignment among crew members, VERY POOR Information provision; 2. Inconsistent work assignment on a daily basis, Unreasonable work assignment among crew members, POOR Information provision; 3. SOMEWHAT Consistent work assignment on a daily basis, Reasonable work assignment among crew members, AVERAGE Information provision; 4. VERY Consistent work assignment on a daily basis, Reasonable work assignment among crew members, GOOD Information provision; 5. VERY Consistent work assignment on a daily basis, Reasonable work assignment among crew members, VERY GOOD Information provision			
1.1.13	Crew participation in foreman decision-making process	Self-explanatory.[13]	Categorical (Decision Type: Without explanation, Joint, and With)	W	SI
1.1.14	Crew flexibility	Refers to the ability and willingness of crew members in performing other member's tasks[14]		D	FM
1.1.14.1	Ability of crew member to perform other's task	Self-explanatory. 100 % percent indicates a crew member is fully capable of handling the other member's task.	Percent (degree of ability to perform other's task).	D	FM
1.1.14.2	Willingness to perform other's tasks	Self-explanatory.	1 - 5 Predetermined rating (shown below)	D	FM
		1. Completely Unwilling; 2. Somewhat NOT Willing; 3. Somewhat Willing; 4. Willing; 5. Completely Willing			
1.1.15	Job site orientation program	Refers to an orientation process to familiarize a crew member with the project, its workers, and any project requirements.	Categorical (Yes, No)	I	SI
1.1.17	Job security	Refers to the level of job security a crew members has in terms of availability of work over the previous year period. Average per crew member.	Integer (Average length of unemployment period, months)	C	SI
1.1.18	Availability of craftsperson	Refers to whether the required number of workers for the activity at hand is met per week.	Integer (Average number of unmet labour demand for the trade under study per week)	W	SI

[12] Sheppard, B. H., and Lewicki, R. J. (1987). "Toward general principles of managerial fairness." *Social Justice Res.*, *1*(2), 161-176.

[13] Heller, F. A., and Yukl, G. (1969). "Participation, managerial decision-making, and situational variables." *Organ. Behav. Human Perfor.*, 4(3), 227-241.

[14] Molleman, E., and van den Beukel, A. (2007). "Worker flexibility and its perceived contribution to performance: The moderating role of task characteristics." *Human Factors Ergonomics in Manuf. Servi. Industries*, 17(2), 117-135.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.1.19	Multiskilling of crew	Refers to the set of skills in terms of trade specialty (carpentry, electrical, ironworker, etc.) the crew members have.	Integer (Total number of trade specialties per crew)	I	SI
1.2.1	Availability of task materials	Refers to whether required task materials are available in site main storage before the commencement of the work.	Real number (Average waiting time for not getting materials at site main storage, min)	W	SI
1.2.2	Quality of task materials	Self-explanatory. It will be measured in terms of level of defect, need for adjustment on site, and level of on-site adjustment.[15]	1 - 5 Predetermined rating (shown below)	W	SI
		1. VERY HIGH Level of defect, YES Need for adjustment on site, VERY FREQUENT Site adjustment; 2. HIGH Level of defect, YES Need for adjustment on site, FREQUENT Site adjustment; 3. MEDIUM Level of defect, YES Need for adjustment on site, FREQUENT Site adjustment; 4. LOW Level of defect, NO Need for adjustment on site; 5. VERY LOW Level of defect, NO Need for adjustment on site			
1.2.3	Availability of temporary material storage	Refers to whether temporary storages are provided near by the location where the task is being executed.	Categorical (Yes, No)	W	SI
1.2.4	Distance to temporary material storage	Self-explanatory.	Real number (Distance, m)	I	DC
1.2.5	Travel time to get materials from site main storage	Refers to the time taken in traveling and getting back with a required material from the site main storage.	Real number (Average travel time for getting materials from main storages, min)	I	DC
1.2.6	Availability of consumables	Refers to whether consumables (nails, duct tapes, drill bits, blades, etc.) are adequately provided.	Categorical (Yes, No)	I	SI
1.2.7	Clear policy on consumables	Refers to whether a clear process is laid out for crew members whereby they can easily and timely get consumables.	Categorical (Yes, No)	I	SI
1.2.8	Material tracking system	Refers to whether an automated material tracking system which tracks material purchases orders, delivery, and follow-up is in place.	Categorical (Yes, No)	I	SI
1.2.9	Material unloading practices	Refers to the time required to unload material from delivery trucks to site lay down area.	Real Number (average unloading time, min)	W	SI
1.2.10	Material movement (horizontal)	Refers to the horizontal distance between the site main storage and the location where the task is being executed.	Real Number (average distance, m)	W	DC
1.2.11	Material movement (vertical)	Refers to the vertical distance between the site main storage and the location where the task is being executed.	Real Number (average distance, m)	W	DC

[15] Ng, S. T., Skitmore, R. M., Lam, K. C., and Poon, A. W. (2004). "Demotivating factors influencing the productivity of civil engineering projects." *Inter. J. Proj. Manage.*, *22*(2), 139-146.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.3.1	Availability of work equipment	Refers to the type and number of work equipment (Zoom boom, Welding machine, Pipe cutter) and the average waiting time in getting required to properly execute the work.16		D	DC
1.3.1.1	Work equipment type and number	Self-explanatory. Each equipment type and the total number of equipment shall be recorded.	Categorical (Type and Number on site)	D	DC
1.3.1.2	Waiting time	Self-explanatory. Waiting time for work equipment.	Real number (Average waiting time, min)	D	FM
1.3.2	Availability of transport equipment	Refers to the type and number of equipment (crane, forklift, pump, crane, bucket, etc.) and the average waiting time in getting required transport workers and material to properly execute the work.		D	FM
1.3.2.1	Transport equipment type and number	Self-explanatory. Each equipment type and the total number of equipment shall be recorded.	Categorical (Type and Number on site)	D	DC
1.3.2.2	Waiting time for transport equipment	Self-explanatory.	Real number (Average waiting time, min)	D	FM
1.3.3	Equipment breakdown	Refers to the recorded breakdown of equipment, for each type of equipment identified in Section 1.3.1.1 and 1.3.2.1	Integer (Equipment Type and Average no. of breakdown occurrence per week)	W	SI
1.3.4	Availability of tools	Refers to the waiting time in getting powered tools (drills, grinders, hammers, etc.), and measured in terms of shortage, misplacement, and sharing of tools.	Real number (Average waiting time, min)	D	SI
1.3.5	Sharing of tools	Refers to the number of crews sharing work tools.	Real number (No. of crews sharing a tool)	W	FM
1.3.6	Quality of tools	Refers to the quality of work tools in terms of number of breakdowns per week.	Integer (Average no. of tool breakdown per week)	W	FM
1.3.7	Efficiency of tool room attendant	Refers to efficiency of tool room attendant in getting tools in a timely fashion.	Real number (Average waiting time for tool, min)	W	SI
1.3.8	Misplacement of tools	Refers to the misplacement of work tools by crew members, which will result in wasted time to locate them.	Real Number (Average no. of misplacement per day)	D	FM
1.3.9	Availability of electric power	Self-explanatory.	Real number (Average waiting time, min)	D	FM
1.3.10	Availability of extension cords	Self-explanatory.	Real number (Average waiting time, min)	D	FM
1.3.11	Quality of tools maintenance	Self-explanatory.	Real number (Average operation time after maintenance, hrs.)	W	SI

16 Turpin, M. P., and Kamath, A. R. R. (1986). "The development of an equipment availability reporting database and analysis package." *Relia. Eng.*,15(2), 95-113.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.4.1	Complexity of task	Refers to the complexity of the task in terms of known alternatives to doing it and the number of subtasks required; measured in terms of unknown or uncertain alternatives, inexact or unknown means - end connections, number of subtasks[17]	1 - 5 Predetermined rating (shown below)	D	SI
		1. MANY Alternatives, WELL KNOWN Means, VERY LOW No. subtasks; 2. SOME Alternatives, WELL KNOWN Means, LOW No. subtasks; 3. FEW Alternatives, KNOWN Means, AVERAGE No. subtasks; 4. FEW Alternatives, UNKNOWN Means, HIGH No. subtasks; 5. VERY FEW Alternatives, UNKNOWN Means, VERY HIGH No. subtasks			
1.4.2	Repetitiveness of task	Refer to how much of the work volume is repetitive in terms of having identical materials and construction methods.[18]	Real number (ratio of identical work tasks quantity to the total work task quantity)	W	SI
1.4.3	Total work volume	Refers to the total quantity approved for construction for the activity under study.	Real number (Approved quantity for construction)	I	SI
1.4.4	Level of Rework	Refers to a work redone for not meeting project requirements in terms of a ratio of the total volume of the rework to the approved total work volume[19]	Percentage (% of activity total volume of rework to total activity work volume)	D	SI
1.4.5	Task change orders - Frequency	Refers to the number of change orders happening in weekly basis.	Real number (No. of occurrence)	W	SI
1.4.6	Task change orders - Extent	Refers to change orders related to the activity under study and their volume in terms of total quantities of work.	Real number (Ratio of approved volume of change order to total work volume)	W	SI
1.4.7	Placement technique	Refers to the method used to execute the work.	Categorical (Pump , Crane and bucket, Direct chute)	D	DC
1.4.8	Building element	Refers to the element of the building/plant under construction	Categorical (Columns, Footings, Grade beams, Pile caps, Slabs, Walls, Boilers)		
1.5.1	Working condition (noise)	Refers to the level of noise: number of equipment creating noise, level of intrusiveness of the created noise, and the felt effect on conversion among workers[20]	1 - 5 Predetermined rating (shown below)	D	FM
		1. NO Noisy Equipment, VERY LOW Intrusiveness, VERY NORMAL Voice Level in Conversation; 2. FEW Noisy Equipment, VERY LOW Intrusiveness, NORMAL Voice Level in Conversation; 3. SOME Noisy Equipment, AVERAGE Intrusiveness, NORMAL Voice Level in Conversation; 4. MANY Noisy Equipment, HIGH Intrusiveness, HIGH Voice Level in Conversation; 5. TOO MANY Noisy Equipment, VERY HIGH Intrusiveness, VERY HIGH Voice Level in Conversation			

[17] Campbell, D.J. (1988). “Task complexity: a review and analysis.” *Acade. Manage. Rev.*, 13(1), 40–52.

[18] COAA (2012). “Benchmarking: Contractor Questionnaire.” *Construction Owners Association Alberta*, Version 8.4, Edmonton, Alberta. Canada.

[19] Fayek, A. Robinson, Dissanayake, M., and Campero, O. (2003). “Measuring and classifying construction field rework: A pilot study.” *Construction Owners Association Alberta*, Field Rework Committee, Edmonton, Alberta, Canada.

[20] HSE (2005). “Noise at work: Guidance for employers on the control of noise at work.” Health and Safety Executive, INDG362, Version 1.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.5.2	Working condition (dust and fumes)	Refers to the level of dust and fumes in work location. It will be measured in terms of the sources, level and length of exposure[21]	1 - 5 Predetermined rating (shown below)	D	FM
		1. NO Source of dust and fume, VERY LOW Level of Exposure, VERY NORMAL Length of Exposure; 2. FEW Sources of dust and fume, VERY LOW Level of Exposure, NORMAL Length of Exposure; 3. SOME Sources of dust and fume, AVERAGE Level of Exposure, NORMAL Length of Exposure; 4. MANY Sources of dust and fume, HIGH Level of Exposure, HIGH Length of Exposure; 5. TOO MANY Sources of dust and fume, VERY HIGH Level of Exposure, VERY HIGH Length of Exposure			
1.5.3	Location of work scope (distance)	Self-explanatory. Distance (horizontal) measured relative to crew's site main trailer.	Real number (distance, m)	D	DC
1.5.4	Location of work scope (elevation)	Self-explanatory. Distance (vertical) measured relative to crew's site main trailer.	Real number (distance, m)	D	DC
1.5.5	Congestion of work area	Refers to the effect of having more workers on the task location than its average (assumed as optimal) number.	Real number (ratio of actual peak manpower to actual average manpower)	W	SI
1.5.6	Cleanliness of work area	Housekeeping to maintain clean work area.	Integer (Number of cleaning operations per day)	D	FM
1.5.7	Cover from weather effect	Refers to whether building envelopes are in place in order to protect workers from weather effects.	Categorical (Yes, No)	I	SI
1.5.8	Location of tool cribs	Location of tool cribs from crew's site main trailer.	Real number (average distance, m)	I	DC
1.5.9	Location of Lunch rooms	Location of lunch room from the crew's site main trailer, in case the lunchroom is not within the crew's site main trailer.	Real number (average distance, m)	I	DC
1.5.10	Location of washrooms	Location of washrooms from the crew's site main trailer. In cases where there is more than one washroom, the average distance is recorded.	Real number (average distance, m)	I	DC
1.6.1	Foreman experience	Refers to the foreman experience in terms of year in industry, after becoming a foreman.	Real number (years of experience)	I	SI
1.6.2	Foreman training	Foremen are expected to get a number of trainings (leadership for safety excellence, CSTS, standard first aid Certificate, supervisory training program) [22] in order to improve their skill. Any training provided to foreman during his/her career, as a foreman is recorded.	Real number (No. trainings attended x Duration of Training, hrs.)	I	SI

[21] HSE(2012). Health and Safety Executive, <http://www.hse.gov.uk/pubns/iacl95.htm> (May 05,2012).

[22] Fayek,A. Robinson, and Poveda, C. (2008). "A pilot study to develop a skills development tool for construction trades foremen." COAA, Supervisory Training and Qualifications, Edmonton, Alberta, Canada.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.6.3	Foreman Skill and Responsibility	Refers to the overall skill sets of the foreman leading the crew.		W	SI
1.6.3.1	Foreman leadership style	Refers to the leadership style of the foreman in terms the following categories: autocratic, democratic, participative, goal-oriented, and situational23	Categorical (Autocratic, Democratic, Participative, Goal-oriented, Situational Definition- Situational)	W	SI
1.6.3.2	Foreman supervision skill	Refers to the supervision skill of the foreman in terms of orientation of crew members; assessing competency and capability of crew members to meet quality requirements; assigning individual and crew tasks; communicating the job to and with the crew; identifying and coordinating job trainings; setting and maintaining work standards to outline behaviour expectation; promoting, supporting and facilitating teamwork and harmony[14]	1 - 5 Predetermined rating (shown below)	W	SI
		1. INADEQUATE in Orientation of crew members; VERY POOR in Assessing competency and capability of crew members to meet quality requirements; VERY POOR in Assigning individual and crew tasks; VERY POOR in Communicating the job to and with the crew; VERY POOR in Identifying and coordinating job training; VERY POOR in Setting and maintaining work standards to outline behaviour expectation; VERY POOR in Promoting, supporting and facilitating teamwork and harmony; 2. INADEQUATE in Orientation of crew members; POOR in Assessing competency and capability of crew members to meet quality requirements; POOR in Assigning individual and crew tasks; POOR in Communicating the job to and with the crew; POOR in Identifying and coordinating job training; POOR in Setting and maintaining work standards to outline behaviour expectation; POOR in Promoting, supporting and facilitating teamwork and harmony; 3. ADEQUATE in Orientation of crew members; FAIR in Assessing competency and capability of crew members to meet quality requirements; FAIR in Assigning individual and crew tasks; FAIR in Communicating the job to and with the crew; FAIR in Identifying and coordinating job training; FAIR in Setting and maintaining work standards to outline behaviour expectation; FAIR in Promoting, supporting and facilitating teamwork and harmony; 4. ADEQUATE in Orientation of crew members; GOOD in Assessing competency and capability of crew members to meet quality requirements; GOOD in Assigning individual and crew tasks; GOOD in Communicating the job to and with the crew; GOOD in Identifying and coordinating job training; GOOD in Setting and maintaining work standards to outline behaviour expectation; GOOD in Promoting, supporting and facilitating teamwork and harmony; 5. ADEQUATE in Orientation of crew members; VERY GOOD in Assessing competency and capability of crew members to meet quality requirements; VERY GOOD in Assigning individual and crew tasks; VERY GOOD in Communicating the job to and with the crew; VERY GOOD in Identifying and coordinating job training; VERY GOOD in Setting and maintaining work standards to outline behaviour expectation; VERY GOOD in Promoting, supporting and facilitating teamwork and harmony			

23 Panthi, K., Farooqui, R. U., and Ahmed, S. M. (2008). "An investigation of the leadership style of construction managers in South Florida." *J. Const. Manage. Econ.*, 11, 455-565.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.6.3.3	Foreman - Provision of clear goals to crafts	Refers to the skill of the foreman in provision and communication of goals[14]	1 - 5 Predetermined rating (shown below)	D	CM
		1. VERY POOR Clarity in assignment of Tasks, VERY POOR Communication; 2. POOR Clarity in assignment of Tasks, POOR Communication; 3. AVERAGE Clarity in assignment of Tasks, AVERAGE Communication; 4. GOOD Clarity in assignment of Tasks, GOOD Communication; 5. VERY GOOD Clarity in assignment of Tasks, VERY GOOD Communication			
1.6.3.4	Foreman - Skill in proper resource allocation	Refers to the skill of the foreman in understanding of schedule and plans, identifying and verifying resource availability, fairness in assignment of resource to different crews, and skill in resolving resource problems[14]	1 - 5 Predetermined rating (shown below)	D	SI
		1. VERY POOR Understanding of schedule & plans, VERY POOR in Identifying & Verifying resource availability, VERY UNFAIR assignment of resource, VERY POOR Skill in resolving resource problems; 2. POOR Understanding of schedule & plans, POOR in Identifying & Verifying resource availability, UNFAIR assignment of resource, POOR Skill in resolving resource problems; 3. FAIR Understanding of schedule & plans, FAIR in Identifying & Verifying resource availability, FAIR assignment of resource, FAIR Skill in resolving resource problems; 4.GOOD Understanding of schedule & plans, GOOD in Identifying & Verifying resource availability, FAIR assignment of resource, GOOD Skill in resolving resource problems; 5. VERY GOOD Understanding of schedule & plans, VERY GOOD in Identifying & Verifying resource availability, VERY FAIR assignment of resource, VERY GOOD Skill in resolving resource problems			
1.6.3.5	Foreman Skill - Safety facilitation and implementation	Refers to the skill of the foreman in safety facilitation and implementation in terms of in knowing, understanding, communicating and ensuring compliance with safety regulation; conducting safety trainings; providing answers to safety related questions; participating and completing safety incident reports[14]	1 - 5 Predetermined rating (shown below)	D	SI
		1. VERY POOR in Knowing, understanding, communicating and ensuring compliance with safety regulation; VERY POOR in Conducting safety trainings; ALWAYS NOT Providing answers to safety related questions; VERY POOR in Participating and completing safety incident reports; 2. POOR in Knowing, understanding, communicating and ensuring compliance with safety regulation; POOR in Conducting safety trainings; SOMETIMES NOT Providing answers to safety related questions; POOR in Participating and completing safety incident reports; 3. FAIR in Knowing, understanding, communicating and ensuring compliance with safety regulation; FAIR in Conducting safety trainings; ADEQUATE in Providing answers to safety related questions; FAIR in Participating and completing safety incident reports; 4. GOOD in Knowing, understanding, communicating and ensuring compliance with safety regulation; GOOD in Conducting safety trainings; ALWAYS in Providing answers to safety related questions; GOOD in Participating and completing safety incident reports; 5. VERY GOOD in Knowing, understanding, communicating and ensuring compliance with safety regulation; VERY GOOD in Conducting safety trainings; ALWAYS in Providing answers to safety related questions; VERY GOOD in Participating and completing safety incident reports			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.6.3.6	Foreman Skill - Planning	Refers to the planning skills of the foreman in terms of identifying and verifying that tools and materials are available and complete; identifying needs and deficiencies in the plan/schedule and communicating these to appropriate persons; translating general work requirements into a prioritized plan for individual tasks and assignments[14]	1 - 5 Predetermined rating (shown below)	D	SI
		1. VERY POOR in Identifying and verifying that tools and materials are available and complete; VERY POOR in Identifying needs and deficiencies in the plan/schedule and communicating these to appropriate persons; VERY POOR in Translating general work requirements into a prioritized plan for individual tasks and assignments; 2. POOR in Identifying and verifying that tools and materials are available and complete; POOR in Identifying needs and deficiencies in the plan/schedule and communicating these to appropriate persons; POOR in Translating general work requirements into a prioritized plan for individual tasks and assignments; 3. FAIR in Identifying and verifying that tools and materials are available and complete; FAIR in Identifying needs and deficiencies in the plan/schedule and communicating these to appropriate persons; FAIR in Translating general work requirements into a prioritized plan for individual tasks and assignments; 4. GOOD in Identifying and verifying that tools and materials are available and complete; GOOD in Identifying needs and deficiencies in the plan/schedule and communicating these to appropriate persons; GOOD in Translating general work requirements into a prioritized plan for individual tasks and assignments; 5. VERY GOOD in Identifying and verifying that tools and materials are available and complete; VERY GOOD in Identifying needs and deficiencies in the plan/schedule and communicating these to appropriate persons; VERY GOOD in Translating general work requirements into a prioritized plan for individual tasks and assignments			
1.6.3.7	Foreman Skill - Scheduling	Refers to the scheduling skills of the foreman in terms of reviewing and adjusting specific workface activities and task schedules to meet established production schedules; working with the crew to overcome work challenges; resolving otherwise reporting scheduling conflicts[14]	1 - 5 Predetermined rating (shown below)	D	SI
		1. VERY POOR in Reviewing and adjusting specific workface activities and task schedules to meet established production schedules; VERY POOR in Working with the crew to overcome work challenges; VERY POOR in Resolving otherwise reporting scheduling conflicts; 2. POOR in Reviewing and adjusting specific workface activities and task schedules to meet established production schedules; POOR in Working with the crew to overcome work challenges; POOR in Resolving otherwise reporting scheduling conflicts; 3. FAIR in Reviewing and adjusting specific workface activities and task schedules to meet established production schedules; FAIR in Working with the crew to overcome work challenges; FAIR in Resolving otherwise reporting scheduling conflicts; 4. GOOD in Reviewing and adjusting specific workface activities and task schedules to meet established production schedules; GOOD in Working with the crew to overcome work challenges; GOOD in Resolving otherwise reporting scheduling conflicts; 5. VERY GOOD in Reviewing and adjusting specific workface activities and task schedules to meet established production schedules; VERY GOOD in Working with the crew to overcome work challenges; VERY GOOD in Resolving otherwise reporting scheduling conflicts			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.6.3.8	Foreman Skill - Employee relation	Refers to the relation of the foreman with crew members in terms of knowing, understanding, communicating and ensuring compliance with all project employee relation requirements; recognizing, addressing and resolving issues/problems among/between crew(s); applying the company's corrective action policy consistently and fairly; applying project procedures, worksite policies and collective agreement requirements[14]	1 - 5 Predetermined rating (shown below)	D	SI
		1. VERY POOR in Knowing, understanding, communicating and ensuring compliance with all project employee relation requirements; VERY POOR in Recognizing, addressing and resolving issues/problems among/between crew(s); VERY POOR in Applying the company's corrective action policy consistently and fairly; VERY POOR in Applying project procedures, worksite policies and collective agreement requirements; 2. POOR in Knowing, understanding, communicating and ensuring compliance with all project employee relation requirements; POOR in Recognizing, addressing and resolving issues/problems among/between crew(s); POOR in Applying the company's corrective action policy consistently and fairly; POOR in Applying project procedures, worksite policies and collective agreement requirements; 3. FAIR in Knowing, understanding, communicating and ensuring compliance with all project employee relation requirements; FAIR in Recognizing, addressing and resolving issues/problems among/between crew(s); FAIR in Applying the company's corrective action policy consistently and fairly; FAIR in Applying project procedures, worksite policies and collective agreement requirements; 4. GOOD in Knowing, understanding, communicating and ensuring compliance with all project employee relation requirements; GOOD in Recognizing, addressing and resolving issues/problems among/between crew(s); GOOD in Applying the company's corrective action policy consistently and fairly; GOOD in Applying project procedures, worksite policies and collective agreement requirements; 5. VERY GOOD in Knowing, understanding, communicating and ensuring compliance with all project employee relation requirements; VERY GOOD in Recognizing, addressing and resolving issues/problems among/between crew(s); VERY GOOD in Applying the company's corrective action policy consistently and fairly; VERY GOOD in Applying project procedures, worksite policies and collective agreement requirements			
1.6.3.9	Foreman Skill - Quality control	Refers to the skill of the foreman in quality control in terms overseeing the execution of the work, by ensuring crew works to job specifications and follows drawings and instructions; inspecting completed work and initiating timely resolutions[14]	1 - 5 Predetermined rating (shown below)	D	SI
		1. VERY POOR in Overseeing the execution of the work, by ensuring crew works to job specifications and follows drawings and instructions; VERY POOR in Inspecting completed work and initiating timely resolutions; 2. POOR in Overseeing the execution of the work, by ensuring crew works to job specifications and follows drawings and instructions; POOR in Inspecting completed work and initiating timely resolutions; 3. FAIR in Overseeing the execution of the work, by ensuring crew works to job specifications and follows drawings and instructions; FAIR in Inspecting completed work and initiating timely resolutions; 4. GOOD in Overseeing the execution of the work, by ensuring crew works to job specifications and follows drawings and instructions; GOOD in Inspecting completed work and initiating timely resolutions; 5. VERY GOOD in Overseeing the execution of the work, by ensuring crew works to job specifications and follows drawings and instructions; VERY GOOD in Inspecting completed work and initiating resolutions			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.6.3.10	Foreman Administration Skill - Supply, Record Keeping	Refers to the administration skill of the foreman in terms of requesting supplies to address any deficiencies in field installation work packages, maintaining foreman's log or diaries, reporting workface production and work progress, completing quality reports, completing required statistics, obtaining permits[16]	1 - 5 Predetermined rating (shown below)	W	SI
		1. VERY POOR in Requesting supplies to address any deficiencies in FIWPs, VERY POOR in Maintaining foreman's log or diaries, Reporting workface production and work progress, Completing quality reports, Completing required statistics, Obtaining permits; 2. POOR in Requesting supplies to address any deficiencies in FIWPs, POOR in Maintaining foreman's log or diaries, Reporting workface production and work progress, Completing quality reports, Completing required statistics, Obtaining permits; 3. FAIR in Requesting supplies to address any deficiencies in FIWPs, FAIR in Maintaining foreman's log or diaries, Reporting workface production and work progress, Completing quality reports, Completing required statistics, Obtaining permits; 4. GOOD in Requesting supplies to address any deficiencies in FIWPs, GOOD in Maintaining foreman's log or diaries, Reporting workface production and work progress, Completing quality reports, Completing required statistics, Obtaining permits; 5. VERY GOOD in Requesting supplies to address any deficiencies in FIWPs, VERY GOOD in Maintaining foreman's log or diaries, Reporting workface production and work progress, Completing quality reports, Completing required statistics, Obtaining permits;			
1.6.3.11	Foreman Administration Skill - Time keeping and personnel	Refers to the administration skill of the foreman in terms of time keeping and time cards (including late starts/early starts); distributing cheques and handling any related issues; recommending personnel actions (hiring, promotions, and discipline)[14]	1 - 5 Predetermined rating (shown below)	W	SI
		1. VERY POOR Time keeping and time cards (including late starts/early starts); VERY POOR in Distributing cheques and handling any related issues; VERY POOR in Recommending personnel actions (hiring, promotions, and discipline); 2. POOR Time keeping and time cards (including late starts/early starts); POOR in Distributing cheques and handling any related issues; POOR in Recommending personnel actions (hiring, promotions, and discipline); 3. FAIR Time keeping and time cards (including late starts/early starts); FAIR in Distributing cheques and handling any related issues; FAIR in Recommending personnel actions (hiring, promotions, and discipline); 4. GOOD Time keeping and time cards (including late starts/early starts); GOOD in Distributing cheques and handling any related issues; GOOD in Recommending personnel actions (hiring, promotions, and discipline); 5. VERY GOOD Time keeping and time cards (including late starts/early starts); VERY GOOD in Distributing cheques and handling any related issues; VERY GOOD in Recommending personnel actions (hiring, promotions, and discipline)			
1.6.3.12	Foreman regard to improvement suggestion by crew members	Self-explanatory.	1 - 5 Predetermined rating (shown below)	W	CM
		1. NO Regard to suggestions; 2. SOME TIMES gives regard to suggestions; 3. OCCASIONALY gives regard to suggestions; 4. COMMONLY gives regard to suggestions; 5. ALWAYS gives regard to suggestions			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.6.3.13	Foreman fairness in performance review of crew by foreman	Self-explanatory.	1 - 5 Predetermined rating (shown below)	W	SI
		1. VERY UNFAIR performance review; 2. UNFAIR performance review; 3. SOMEWHAT FAIR performance review; 4. FAIR performance review; 5. VERY FAIR performance review			
1.6.4	Change of foremen	Refers to whether there is a change in foreman.	Turnover rate (No. of turnovers per month)	M	SI
1.6.5	Span of control	Refers to the total number of crews controlled by the foreman.	Integer (total number of crews per foreman)	W	SI
1.6.6	Use of assistant foremen	Self-explanatory.	Categorical (Yes, No)	I	SI
1.6.7	Provision of feedback on foreman's performance	Refers to provision of feedback to foreman by supervisors or project managers.	Categorical (Yes, No)	I	SI
1.7.1	Availability of drawings	Refers to whether required work drawings are found available on site.	1 - 5 Predetermined rating (shown below)	M	SI
		1. Always Not Available; 2. Sometimes Not Available; 3. Sometimes Available; 4. Mostly Available; 5. Always Available			
1.7.2	Quality of drawings	Refers to the quality of the drawings in terms of completeness, readability, reusability, clarity of information, and frequency of updates.	1 - 5 Predetermined rating (shown below)	M	SI
		1. Incomplete, VERY POOR Readability, VERY LOW Reusability, TOO MANY Unclear information, NOT Updated; 2. Incomplete, POOR Readability, LOW Reusability, SOME Unclear informations, NOT Updated; 3. Incomplete, AVERGAE Readability, AVERAGE Reusability, FEW Unclear informations, NOT Updated; 4. Complete, GOOD Readability, HIGH Reusability, FEW Unclear informations, Updated; 5. Complete, VERY GOOD Readability, HIGH Reusability, VERY FEW Unclear informations, Updated			
1.7.3	Number of drawing revisions	Number of drawing revisions submitted to site foreman per week. Drawing revisions shall be approved and submitted to crew members by foreman.	Integer (Number of drawing revision per week)	W	SI
1.7.4	Specifications			I	
1.7.4.1	Use of standard specifications	Using standard specification like MasterFormat, UNIFORMAT, or others.	Categorical (Yes, No)	I	SI
1.7.4.2	Availability of specifications	Refers to whether required work specifications are found available on site.	1 - 5 Predetermined rating (shown below)	I	SI
		1. Always Not Available; 2. Sometimes Not Available; 3. Sometimes Available; 4. Mostly Available; 5. Always Available			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
1.7.4.3	Quality of specification	Refers to the quality of the specifications in terms of completeness, and clarity of information. Measured in terms of clarity and completeness.	1 - 5 Predetermined rating (shown below)	I	SI
		1. VERY POOR Clarity, VERY Incomplete; 2. POOR Clarity, Incomplete; 3. FAIR Clarity, FAIRLY Complete; 4. GOOD Clarity, Complete; 5. VERY GOOD Clarity, VERY Complete			
1.7.4.4	Number of specification revisions	Number of specification revisions submitted to site foreman per week. Specification revisions shall be approved and submitted to crew members by foreman.	Integer (Number of specification revision per week)	W	SI
1.7.6	Response rate with RFI's	Refers to the response time to request for information (RFI) from the contractor to owner and/or engineer.	Real number (Average response time, hrs.)	W	SI
1.7.7	Adequacy of instructions	Refers to the adequacy of work instructions in terms of information on work procedure, construction steps, and communication means to crew members from foreman/superintendent24	1 - 5 Predetermined rating (shown below)	I	CM
		1. NO Information on work procedure, NO Information on construction steps, Communication means not clearly laid out; 2. SOME Information on work procedure, SOME Information on construction steps, Communication means not clearly laid out; 3.ADEQUATE Information on work procedure, ADEQUATE Information on construction steps, Communication means SOMEWHAT laid out; 4. VERY GOOD Information on work procedure, VERY GOOD Information on construction steps, Communication means laid out; 5. EXCELLENT Information on work procedure, EXCELLENT Information on construction steps, Communication means clearly laid out			

24 Mourgues, C, and Fischer, M. (2008). "A work instruction template for cast-in-place concrete construction laborers." *CIFE*, Working paper, 109, Stanford Univ., California, US.

A.3: Quantification of Project Level Input Parameters

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.1.1	Project delivery system	Refers to the arrangement between the owner and contractor on the means to design, execute, and operate the project.	Categorical (Design bid build, Design build, Build operate transfer, Private public partnership)	I	SI
2.1.2	Contract type	Refers to the contract arrangement made for the project between the owner and contractor or general contractor and subcontractor.	Categorical (Lump sum, Unit rate, Cost reimbursable)	I	PM
2.1.3	Level of fast tracking	Refers to whether the project construction begun before the completion of the design process.	Real number (% Overlap between design and construction schedule)	I	PM
2.1.4	Change in design drawings	Number of drawing revisions submitted at a project level for all activities.	Real number (Ratio of number of changed drawings to total number of drawings per project)	M	SI
2.1.5	Change in specifications	Number of specification revisions submitted at a project level for all activities.	Real number (Ratio of number of changed specifications to total number of specification clauses at the project level)	M	SI
2.1.6	Changes in contract conditions	Number of revisions on contract conditions submitted at a project level for all activities.	Real number (Ratio of number of contract conditions changes to total number of contract clauses at the project level)	M	PM
2.1.7	Lack of information	Refers to the lack of information associated with the design and execution of the project and will be measured in terms of request for information (RFI) per month.	Real number (Number of RFI's per month per discipline)	M	PM
2.1.8	Approval for building permit	Refers to the time taken to get the permit to build from appropriate municipality offices.	Real number (average process time for work or permit approval, months)	M	PM
2.2.1	Project type	Self-explanatory.	Categorical (Commercial, Institutional, Residential, Industrial)	I	DC
2.2.2	Project size	Refers to the project size in terms of contract dollar value.	Real number (Project contract value,$)	I	PM
2.2.3	Project complexity	Refers to the overall complexity of the projects in terms of use of unproven technology, number of process steps, facility size or process capacity, past experience with configuration or geometry, and construction methods25		I	PM
2.2.3.1	Extent of use of new and unproven technology	Self-explanatory.	Categorical (Yes, No)	I	PM

[25] COAA (2012). "Benchmarking: Contractor Questionnaire." *Construction Owners Association Alberta*, Version 8.4, Edmonton, Alberta, Canada.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.2.3.2	Extent of use of technological advanced methods	Self-explanatory.	1 - 5 Predetermined rating (shown below)	I	PM
		1. Very Low; 2. Low; 3. Average; 4. High; 5. Very High			
2.2.3.3	Construction methods and techniques	Refers to the construction methods and techniques adopted for the project and the experience and availability of proper procedure with methods and technologies.	1 - 5 Predetermined rating (shown below)	I	PM
		1. VERY POOR Experience with methods and technology, LACK of proper procedure; 2. POOR Experience with methods and technology, LACK of proper procedure; 3. FAIR Experience with methods and technology, WITH proper procedure; 4. GOOD Experience with methods and technology, WITH proper procedure; 5. VERY GOOD Experience with methods and technology , WITH proper procedure			
2.2.3.4	Facility size or process capacity	Refers to the size of the project in terms of know measures like total built floor space, total man-hours, etc.	Real number (Total unit)	I	PM
2.2.3.5	Past experience with configuration or geometry	Self-explanatory.	Real number (Total number of similar projects completed)	I	PM
2.2.4	Project location	Refers to the city where the project is being implemented.	Categorical (Edmonton, Nisku, Acheson, Calgary, Fort McMurray, Other)	I	DC
2.2.5	Year of construction	Refers to the year the data collection of the project began.	Integer (Year of Construction)	M	DC
2.2.6	Amount of modularization	Refers to the amount of modularization in terms of off-site work in a module yard or prefabrication plant.	Real number (% off site construction cost to total project cost)	I	PM
2.2.7	Project organization structure	Refers to the way the project team is set.	Categorical (Line, Product, Functional, Matrix)	I	DC
2.2.8	Project level rework	Refers to the total amount of rework at the project level, including all activities. Project Construction Filed Rework Index CFRI (% of Total Cost of rework to total field construction phase cost)26	Real number (Project overall CFRI)	M	PM
2.2.9	Project level change order	Refers to the total amount of change order at the project level, including all activities.	Real number (% approved total cost of change order overall project to original approved project cost)	M	PM
2.2.10	Project percent complete	Refers to how much of the project is completed in terms of construction project cost.	Real number (% complete of approved construction project cost)	I	PM

[26] Fayek, A. Robinson, Dissanayake, M., and Campero, O. (2003). “Measuring and classifying construction field rework: A pilot study.” *Construction Owners Association Alberta*, Field Rework Committee, Edmonton, Alberta, Canada.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.3.1	Site transportation	Self-explanatory.		I	DC
2.3.1.1	Flight arrangements	Refers to whether the site is remote enough that flight arrangements are made.	Categorical (Yes, No)	I	DC
2.3.1.2	Provision of ground transportation for workers to site	Self-explanatory.	Categorical (Yes, No)	I	DC
2.3.2	Camp condition	Refers to whether the site is remote enough that camp facilities are provided. The condition of the camp will be evaluated in terms of room condition, food service, and amenities.	1 - 5 Predetermined rating (shown below)	I	SI
		1. VERY POOR Room Condition, POOR Food service, NO Amenities; 2. POOR Room Condition, POOR Food service, NO Amenities; 3. FAIR Room Condition, FAIR Food service, SOME Amenities; 4. GOOD Room Condition, GOOD Food service, SOME Amenities; 5. VERY GOOD Room Condition, VERY GOOD Food service, MANY Amenities			
2.3.3	Weather (temperature)	Refers to the recorded temperature at 1:00 PM of the work day.	Real number (˚C)	D	DC
2.3.4	Weather (precipitation)	Refers to the recorded daily average precipitation of the work day.	Real number (mm)	D	DC
2.3.5	Weather (humidity)	Refers to the recorded daily average humidity of the work day.	Real number (%)	D	DC
2.3.6	Weather (wind speed)	Refers to the recorded daily average wind speed of the work day.	Real number (km/hr)	D	DC
2.3.7	Weather (radiation)	Refers to the highest recorded radiation of the work day.	Real number (Hz)	W	DC
2.3.8	Variability of weather	Refers to the variability of the weather on a weekly basis. Average number of heating degree days (HDD), number of cold degree days (CDD), total precipitation per project area, and standard deviation of precipitation[27]		W	DC
2.3.8.1	Number of heating degree days (HDD)	Heating degree-days for a given day are the number of degrees Celsius that the mean temperature is below 18°C.	Integer (Number of days per week)	W	DC
2.3.8.2	Number of cold degree days (CDD)	Cooling degree-days for a given day are the number of degrees Celsius that the mean temperature is above 18°C.	Integer (Number of days per week)	W	DC
2.3.9	Ground conditions	Refers to the trafficability of the ground for walking around during work times.	1 - 5 Predetermined rating (shown below)	W	SI
		1. HIGH Moisture content, Clay Soil; 2. MEDIUM Moisture content, Clay Soil; 3. MEDIUM Moisture content, Sandy Clay; 4. MEDIUM Moisture content, Sandy Soil; 5. Low Moisture content, Sandy Soil			

27 Moosavi, S.F., and Moselhi,O. (2012). “Schedule assessment and evaluation.” *Proc., Construction Research Congress*, ASCE, West Lafayette, Indiana, US, 535 – 544.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.3.10	Site layout	Refers to the efficiency of the site layout in terms of temporary facilities (laydown area, warehouse, fabrication shops, and batch plant). It will be measured in terms of identification of facilities, efficiency of placement of the different areas, and size requirements.	1 - 5 Predetermined rating (shown below)	I	SI
		1. VERY POOR Identification, POOR Placement, VERY LARGE Size requirement; 2. POOR Identification, POOR Placement, LARGE Size requirement; 3. GOOD Identification, POOR Placement, LARGE Size requirement; 4. GOOD Identification, GOOD Placement, AVERAGE Size requirement; 5. VERY GOOD Identification, VERY GOOD Placement, SMALL Size requirement			
2.3.11	Site congestion	Refers to the congestion of site, measured in terms of available free site space.	Real number (Ratio free site space to total site area)	W	SI
2.3.12	Site access	Refers to the efficiency of site access to project facility location in terms of crowding, width of access, and waiting time.		W	DC
2.3.12.1	Width of access	Self-explanatory.	Real number (Width of access, m)	W	DC
2.3.12.2	Queue time to access site	Self-explanatory.	Real number (Average queue time to access time, minutes)	W	DC
2.3.13	Parking facilities (within project)	Self-explanatory.	Integer (Ratio parking capacity of plot to total number of workers)	I	DC
2.3.14	Site storage	Refers to the total on-site storage area available.	Real number (Ratio of total on-site storage area to total site area)	I	DC
2.3.15	Site facilities for workers	Self-explanatory.		I	DC
2.3.15.1	Site facilities for workers (lunch room)	Self-explanatory.	Real number (Average size, m^2 and number of lunch room on project)	I	DC
2.3.15.2	Site facilities for workers (wash room)	Self-explanatory.	Integer (Average size, m^2 and number of wash rooms on project)	I	DC
2.3.16	Unloading/laydown area	Refers to the size of the site's unloading/laydown area.	Real number (Ratio total area of unloading/laydown to total project site area)	I	DC
2.3.17	Project work times	Refers to the work times in terms of shift length, use of multiple shifts, and use of overtime.		W	SI
2.3.17.1	Use of overtime	Self-explanatory. Overall project activities (besides the one under study).	Real number (Average over time per week)	W	SI
2.3.17.2	Multiple shifts (night)	Self-explanatory. Use of night shifts in project including three shifts	Categorical (Yes, No)	W	SI

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.3.17.3	Shift length	Self-explanatory. Excessiveness of shift lengths.	Real number (hours per shift)	W	SI
2.4.1	Owner staff on site	Refers to the total number of owner staff on site to supervise the project works.	Integer (Total number of owner staff on site)	I	PM
2.4.2	Supervision from owner	Refers to the level of supervision by owner representatives.	1 - 5 Predetermined rating (shown below)	I	PM
		1. NO Supervision; 2. VERY SMALL Supervision; 3. SOME Supervision; 4. HIGH Supervision, 5. VERY HIGH Supervision			
2.4.3	Owner's primary driver	Refers to the primary focus of the owner among the project performance objectives.	Categorical (Schedule, Cost, Quality, Safety)	I	PM
2.4.4	Delivery of site to contactor	Refers to the days taken to handover the site after notice to proceed.	Real number (Days taken to handover site, days)	I	PM
2.4.5	Approval of shop drawings and sample materials	Refers to the time taken in approving shop drawings and sample materials by the owner or his representatives.	Real number (Average time taken to approve, days)	M	PM
2.4.6	Suspension of project work (owner reasons)	Refers to whether the project was suspended during the execution process.		I	PM
2.4.6.1.	Number of suspensions	Refers if the project has been suspended during its execution process.	Integer (Number of suspensions due to owner)	I	PM
2.4.6.2	Length of suspensions	Self-explanatory.	Real number (Average length of suspensions, days)	I	PM
2.5.1	Experience of project management team members	Refers to the experience of the contractor project staff in terms of years in industry.	Real number (Average years of experience of PM team)	I	PM
2.5.2	Support and administrative staff	Refers to number of support and administrative staff (secretary, drivers, and tool crib attendants) on site.	Real number (Ratio of support to technical staff)	M	PM
2.5.3	Level of paper work for work approval	Refers to the level of paper work required to get work approvals from owner or his representatives.		W	PM
2.5.3.1	Forms	Self-explanatory.	Integer (number of forms to be filled)	W	PM
2.5.3.2	Approval Signatures	Self-explanatory.	Integer (number of approval signatures)	W	PM
2.5.4	Treatment of foremen by superintendent and project manager	Self-explanatory.	1 - 5 Predetermined rating (shown below)	W	PM
		1. ALWAYS Disrespectful, Insincere, NO Counselling; 2. OFTEN Disrespectful, Insincere, NO Counselling; 3. SOMETIMES Respectful, Sincere, Counselling; 4. OFTEN Respectful, Sincere, Counselling; 5. ALWAYS Respectful, Sincere, Counselling			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.5.5	Performance competition system within the company	Refers to whether a performance based competition system for trade workers exists within the company.	Categorical (Yes, No)	I	PM
2.5.6	Uniformity of work rules by superintendent	Self-explanatory.	1 - 5 Predetermined rating (shown below)	W	PM
		1. VERY Irregular among crews and HIGHLY Variable in daily work times and work days; 2. Irregular among crews and Variable in daily work times and work days; 3. Uniform among crews and Variable in daily work times and work days; 4. Uniform among crews, Always the same in daily work times and work days; 5. VERY Uniform among crews, Always the same in daily work times and work days			
2.5.8	Superintendent education	Self-explanatory.	Categorical (Below Secondary, Secondary School, Technical or Apprentice, College, University)	I	SI
2.5.9	Superintendent training	Self-explanatory. Trainings related to time management, leadership for safety excellence, CSTS, standard first aid, supervisory training program are documented[28]	Integer (No. trainings attended x Duration of Training)	I	SI
2.5.11	Project Manager education	Self-explanatory.	Categorical (Below Secondary, Secondary School, Technical or Apprentice, College, University)	I	PM
2.5.12	Project manager training	Self-explanatory. Trainings related to time management, cost, quality, certificate for Project Management Professional (PMP) are documented.	Real number (No. trainings attended x Duration of Training)	I	PM
2.6.1	Labour union type	Self-explanatory.	Categorical (Building Trades, CLAC, Non-union)	I	DC
2.6.2	Availability of labour	Refers to whether the required number of workers for the all activities in the project are met per month.	Integer (Unmet labour requirement, for the given trade)	M	SI

[28] Fayek,A. Robinson, and Poveda, C. (2008). “A pilot study to develop a skills development tool for construction trades foremen.” *COAA*, Supervisory Training and Qualifications, Edmonton, Alberta, Canada.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.6.3	Labour Disputes (legal cases between a worker on a project)	Self-explanatory.	Integer (Average number of cases per project)	M	SI
2.6.4	Union influences	Self-explanatory.	1 - 5 Predetermined rating (shown below)	I	SI
		1. NO Influence; 2. VERY SMALL Influence; 3. SOME Influence; 4. HIGH Influence, 5. VERY HIGH Influence			
2.7.1	Salary (Project Manager)	Self-explanatory.	Real number (Average annual salary, thousands)	I	DC
2.7.2	Salary (Superintendent)	Self-explanatory.	Real number (Average annual salary, thousands)	I	DC
2.7.3	Salary (Foreman)	Self-explanatory.	Real number (Average annual salary, thousands)	I	DC
2.7.4	Salary (Craftsperson, Journeyman)	Self-explanatory.	Real number (Average annual salary, thousands)	I	DC
2.7.5	Salary (Craftsperson, Apprentice)	Self-explanatory.	Real number (Average annual salary, thousands)	I	DC
2.7.6	Benefits	Refers to benefits (medical and other) provided to workers.	Real number (Average benefits for a craftsperson per day)	M	DC
2.8.1	Detailed front end planning	Refers to efficiency of the detailed front end planning of the project, measured in terms of team composition, technology use in evaluation (like Simulation), evaluation for alternate options, level of risk analysis[29]	1 - 5 Predetermined rating (shown below)	I	PM
		1. INEXPERIANCED Team, NO Use of technological methods, ONLY FEW Alternative evaluated, NO Risk analysis; 2. INEXPERIANCED Team, NO Use of technological methods, FEW Alternative evaluated, NO Risk analysis; 3. EXPERIANCED Team, NO Use of technological methods, SOME Alternatives evaluated, SOME form of Risk analysis; 4. EXPERIANCED Team, SOME Use of technological methods, SOME Alternatives evaluated, SOME form of Risk analysis; 5. WELL EXPERIANCED Team, DETAILED Use of technological methods, MANY Alternatives evaluated, DETAILED Risk analysis			

[29] COAA (2012). “Benchmarking: Contractor Questionnaire.” *Construction Owners Association Alberta*, Version 8.4, Edmonton, Alberta, Canada.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.8.2	Constructability Review	Refers to the efficiency of constructability reviews between owner and contractor representatives.		I	PM
2.8.2.1	Defining project objectives for constructability review	Self-explanatory.	1 - 5 Predetermined rating (shown below)	I	PM
		1. Project Objective/s NOT identified, Identified project objective/s NOT properly implemented; 2. Project Objective/s SOMEWHAT identified, Identified project objective/s SOMEWHAT implemented; 3. Project Objective/s identified, Identified project objective/s SOMEWHAT implemented; 4. Project Objective/s identified, Identified project objective/s SOMEWHAT Properly implemented; 5. Project Objective/s identified, Identified project objective/s Properly implemented			
2.8.2.2	Constructability objective and measure	Self-explanatory.	1 - 5 Predetermined rating (shown below)	I	PM
		1. Not Well Defined Objectives, Constructability improvements VERY POORELY measured; 2. Not Well Defined Objectives, Constructability improvements POORELY measured; 3. FAIRLY Defined Objectives, Constructability improvements POORELY measured; 4. Well Defined Objectives, Constructability improvements measured; 5. Very Well Defined Objectives, Constructability improvements measured to detail			
2.8.2.3	Constructability ideas and implementation	Self-explanatory.		I	PM
A	Constructability Ideas	Self-explanatory.	Integer (Number of well-defined ideas)	I	PM
B	Implementation of constructability ideas	Self-explanatory.	Categorical (Yes, No)	I	PM
2.9.1	Project scope definition	The process of subdividing the major project deliverables into smaller, more manageable components to develop a project Work Breakdown Structure (WBS)	1 - 5 Predetermined rating (shown below)	I	PM
		1. Defined project scope NOT properly used to define project WBS, VERY POOR Experience in work decomposition, Developed WBS NOT Comprehensively covering the project scope; 2. Defined project scope NOT properly used to define project WBS, POOR Experience in work decomposition, Developed WBS NOT Comprehensively covering the project scope; 3. Defined project scope properly used to define project WBS, FAIR Experience in work decomposition, Developed WBS NOT Comprehensively covering the project scope; 4. Defined project scope properly used to define project WBS, GOOD Experience in work decomposition, Developed WBS SOMEHOW Comprehensively covering the project scope; 5. Defined project scope properly used to define project WBS, VERY GOOD Experience in work decomposition, Developed WBS Comprehensively covering the project scope			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.9.2	Project scope verification	The process of obtaining formal acceptance of the project scope by the stakeholders	1 - 5 Predetermined rating (shown below)	I	PM
		1. Project scope verification NOT conducted; 2. Project scope verification SOMEWHAT conducted; 3. Project scope verification PARTIALLY conducted; 4. Project scope verification MOSTLY conducted; 5. Project scope verification FULLY conducted			
2.9.3	Project scope change control	Deals with influencing the factors causing scope changes, determining that a scope change has occurred and managing actual changes when and if they occur.	1 - 5 Predetermined rating (shown below)	I	PM
		1. LACK of project change documents, NO procedure for change management tracking and approval, VERY POOR performance measurement system, VERY POOR Integration with other control processes; 2. LACK of project change documents, NO procedure for change management tracking and approval, POOR performance measurement system, POOR Integration with other control processes; 3. PRESENCE of project change documents, NO procedure for change management tracking and approval, FAIR performance measurement system, FAIR Integration with other control processes; 4. PRESENCE of project change documents, EXISTING procedure for change management tracking and approval, GOOD performance measurement system, GOOD Integration with other control processes; 5. PRESENCE of project change documents, EXISTING procedure for change management tracking and approval, VERY GOOD performance measurement system, VERY GOOD Integration with other control processes			
2.10.1	Project planning and scheduling	Self-explanatory. Contractual compliance (milestones, scope coverage, activity duration), Schedule development (Scope definition, WBS, Scheduling participation, subcontractor participation), Schedule components (Job logic, critical paths, special consideration)[30]		I	PM
2.10.1.1	Project activity definition	Involves identifying and documenting the specific activities that must be performed to produce the deliverables identified in the WBS	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Use of project information (WBS, Scope statement), NOT properly documentation assumptions, VERY POORELY decomposing activities from WBS, NOT Using concurrent engineering ideas; 2. POOR Use of project information (WBS, Scope statement), NOT properly documentation assumptions, POORELY decomposing activities from WBS, NOT Using concurrent engineering ideas; 3. AVERAGE Use of project information (WBS, Scope statement), NOT properly documentation assumptions, FAILRY decomposing activities from WBS, NOT Using concurrent engineering ideas; 4. GOOD Use of project information (WBS, Scope statement), Properly documentation assumptions, GOOD in decomposing activities from WBS, Using SOME concurrent engineering ideas; 5. VERY GOOD Use of project information (WBS, Scope statement), Properly documentation assumptions, VERY GOOD in decomposing activities from WBS, Using many concurrent engineering ideas			

[30] Moosavi, S.F., and Moselhi,O. (2012). “Schedule assessment and evaluation.” *Proc., Construction Research Congress*, ASCE, West Lafayette, Indiana, US, 535 – 544.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.10.1.2	Project activity sequencing	Involves identifying and documenting interactivity logical relationships	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Understanding of technical and resource dependencies between activities, NOT Using activity sequencing tools; 2. POOR Understanding of technical and resource dependencies between activities, NOT Using activity sequencing tools; 3. FAIR Understanding of technical and resource dependencies between activities, FAIR Use of activity sequencing tools; 4. GOOD Understanding of technical and resource dependencies between activities, GOOD Use of activity sequencing tools; 5. VERY GOOD Understanding of technical and resource dependencies between activities, VERY GOOD Use of activity sequencing tools			
2.10.2	Project activity duration	Self-explanatory.		W	SI
2.10.2.1	Project activity duration estimation	The process of taking information on project scope and resources and then developing durations for input to schedules	1 - 5 Predetermined rating	W	SI
		1. Development of resource requirements and resource capabilities NOT PROPERLY done, NO Use of historical information (past project files, commercial databases like RS Means), Experience of estimator VERY POOR; 2. Development of resource requirements and resource capabilities NOT PROPERLY done, NO Use of historical information (past project files, commercial databases like RS Means), Experience of estimator POOR; 3. Development of resource requirements and resource capabilities SOMEWHAT done, SOME Use of historical information (past project files, commercial databases like RS Means), Experience of estimator FAIR; 4. Development of resource requirements and resource capabilities WELL done, AVERAGE Use of historical information (past project files, commercial databases like RS Means), Experience of estimator GOOD; 5. Development of resource requirements and resource capabilities VERY WELL done, EXCELLENT Use of historical information (past project files, commercial databases like RS Means), Experience of estimator VERY GOOD			
2.10.2.2	Unrealistic activity duration	Self-explanatory.	1 - 5 Predetermined rating	W	SI
		1. VERY Unrealistic; 2. Unrealistic; 3. Common industry average; 4. Realistic; 5. VERY Realistic			
2.10.3	Project schedule development	Develops the start and finish dates for project activities	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Understanding of constraints, project calendar, resource plans, VERY POORLEY Developed activity attributes, NOT Using scheduling tools (CPM, PERT, Simulation), NO Use of resource leveling and Project management software's; 2. POOR Understanding of constraints, project calendar, resource plans, POORLEY Developed activity attributes, NOT Using scheduling tools (CPM, PERT, Simulation), NO Use of resource leveling and Project management software's; 3. FAIR Understanding of constraints, project calendar, resource plans, FAIRLY Developed activity attributes, FAIR Use scheduling tools (CPM, PERT, Simulation), FAIR Use of resource leveling and Project management software's; 4. GOOD Understanding of constraints, project calendar, resource plans, WELL Developed activity attributes, GOOD Use scheduling tools (CPM, PERT, Simulation), GOOD Use of resource leveling and Project management software's; 5. VERY GOOD Understanding of constraints, project calendar, resource plans, WELL Developed activity attributes, VERY GOOD Use scheduling tools (CPM, PERT, Simulation), VERY GOOD Use of resource leveling and Project management software's			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.10.3.1	Project duration	Refers to whether the contractual project duration is realistic or not.	1 - 5 Predetermined rating	I	PM
		1. VERY Unrealistic; 2. Unrealistic; 3. Common industry average; 4. Realistic; 5. VERY Realistic			
2.10.3.2	Criticality of project schedule	Self-explanatory.	Real number (ratio of critical to total number of activities)	I	PM
2.10.4	Project schedule control	Deals with influencing the factors causing schedule changes, determining that a schedule change has occurred and managing actual changes when and if they occur		I	PM
2.10.4.1	Schedule Compression	Refers to whether any of the activities have been crushed to achieve saving in project duration.	Real number (Ratio crushed to normal schedule)	M	PM
2.10.5	Project activity weights definition	Involves evaluating activities characteristics and attributes in order to assess the contribution of each particular project activity to the overall project progress of a given phase or deliverable of the project	1 - 5 Predetermined rating	I	PM
		1. Defining activity attributes in terms of durations, costs, labour hours, quantities NOT DONE, NO Use of expert judgment, NO Use of percentage calculation; 2. Defining activity attributes in terms of durations, costs, labour hours, quantities SOMEWHAT, NO Use of expert judgment, NO Use of percentage calculation; 3. Defining activity attributes in terms of durations, costs, labour hours, quantities PARTIALLY DONE, NO Use of expert judgment, YES to Use of percentage calculation; 4. Defining activity attributes in terms of durations, costs, labour hours, quantities MOSTLY DONE, YES Use of expert judgment, YES Use of percentage calculation; 5. Defining activity attributes in terms of durations, costs, labour hours, quantities FULLY DONE, YES Use of expert judgment, YES Use of percentage calculation			
2.10.6	Project progress curves development and Progress monitoring	Refers to the creation of a progress baseline to monitor actual versus planned performance	1 - 5 Predetermined rating	I	PM
		1. NO Use of project schedules, activity weights, Standard Performance curves NOT developed, NO Use of project management software; 2. POOR Use of project schedules, activity weights, Standard Performance curves NOT developed, NO Use of project management software; 3. FAIR Use of project schedules, activity weights, Standard Performance curves POORLY developed, NO Use of project management software; 4. GOOD Use of project schedules, activity weights, Standard Performance curves WELL developed, Use of project management software; 5. VERY GOOD Use of project schedules, activity weights, Standard Performance curves VERY WELL developed, Use of project management software			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.11.1	Project resource planning	Determining what resources (people, equipment, materials) and what quantities of each should be used to perform project activities. Measured using use of project information (Scope, WBS, Historical information, Contract requirement, Activity duration), understanding of construction methods, use of Project management software's, and development of detailed resource requirement with a resource profile and schedule	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Use of project information (WBS, Scope statement, Activity duration, Historical records), NO Use of Project management software's, INADEQUATELY Developed resource plan; 2. POOR Use of project information (WBS, Scope statement, Activity duration, Historical records), NO Use of Project management software's, INADEQUATELY Developed resource plan; 3. FAIR Use of project information (WBS, Scope statement, Activity duration, Historical records), NO Use of Project management software's, INADEQUATELY Developed resource plan; 4. GOOD Use of project information (WBS, Scope statement, Activity duration, Historical records), Use of Project management software's, ADEQUATELY Developed resource plan; 5. VERY GOOD Use of project information (WBS, Scope statement, Activity duration, Historical records), ADVANCED Use of Project management software's, ADEQUATELY Developed resource plan			
2.11.2	Project cost estimating	Developing an approximation (estimate) of the costs of the resources needed to complete project activities. Measured using basic estimation process details (developing material & equipment list, project schedule), team experience, time allowed for estimation, bidding and labour climate[31]		I	PM
2.11.2.1	Development of material, equipment list	Self-explanatory.	1 - 5 Predetermined rating	I	PM
		1. VERY POORLY Developed material and equipment list; 2. POORLY Developed material and equipment list; 3. FAIRLY Developed material and equipment list; 4. WELL DONE material and equipment list; 5. VERY WELL DONE material and equipment list			
2.11.2.2	Estimation team experience	Self-explanatory.	Real number (Average years of experience of estimation team)	I	PM
2.11.2.3	Time allowed for estimation	Self-explanatory.	Integer (Time taken for estimation, working days)	I	PM

[31] Trost, S., and Oberlender,G. (2003). "Predicting accuracy of early cost estimates using factor analysis and multivariate regression." *J. Constr. Eng. Manage.*, 129(2), 198 – 204.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.11.2.4	Bidding climate	Refers to the bidding climate in terms of uncertainty in future, quality of bid document, competition level, and project type. Uncertainty in future, Quality of bid document, Competition level, Type of project	1 - 5 Predetermined rating	I	PM
		1. VERY HIGH Uncertainty in future, VERY POOR Quality of bid document, VERY HIGH Competition level, UNFAVOURABLE Type of project; 2. HIGH Uncertainty in future, POOR Quality of bid document, HIGH Competition level, UNFAVOURABLE Type of project; 3. FAIR Uncertainty in future, AVERAGE Quality of bid document, MEDIUM Competition level, FAVOURABLE Type of project; 4. FAIR Uncertainty in future, GOOD Quality of bid document, LOW Competition level, FAVOURABLE Type of project; 5. LOW Uncertainty in future, VERY GOOD Quality of bid document, VERY LOW Competition level, FAVOURABLE Type of project			
2.11.2.5	Labour climate	Refers to the labour climate during bidding and estimating, measured in terms of availability of labour, quality of labour, and agreements with Unions	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Availability of labour, VERY POOR Quality of labour, NO Agreement with Unions; 2. POOR Availability of labour, POOR Quality of labour, NO Agreement with Unions; 3. FAIR Availability of labour, FAIR Quality of labour, YES Agreement with Unions; 4. GOOD Availability of labour, GOOD Quality of labour, YES Agreement with Unions; 5. VERY GOOD Availability of labour, VERY GOOD Quality of labour, YES Agreement with Unions			
2.11.3	Project cost budgeting	Involves allocating the overall cost estimate to individual work activities or work packages to establish a cost baseline for measuring project cost performance. Use of project information (Cost estimates, WBS, project schedule), Use of cost budgeting tools and techniques (Computerized tools), Development of a cost baseline (a time-phased budget to be used for measuring and monitoring cost performance of project)	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Use of project information (Cost estimates, WBS, project schedule), NO Use of computerized tools, INADEQUATELY Developed cost baseline; 2. POOR Use of project information (Cost estimates, WBS, project schedule), NO Use of computerized tools, INADEQUATELY Developed cost baseline; 3. FAIR Use of project information (Cost estimates, WBS, project schedule), SOME Use of computerized tools, INADEQUATELY Developed cost baseline; 4. GOOD Use of project information (Cost estimates, WBS, project schedule), Use of computerized tools, ADEQUATELY Developed cost baseline; 5. VERY GOOD Use of project information (Cost estimates, WBS, project schedule), ADVANCED Use of computerized tools, ADEQUATELY Developed cost baseline			
2.11.4	Project cost control	Deals with influencing the factors causing changes to cost baseline, determining that the cost baseline has changed and managing actual changes when and if they occur.		M	PM

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.11.4.1	Use of Earned value methods	Self-explanatory.	1 - 5 Predetermined rating	M	PM
		1. Earned value methods NOT employed; 2. Earned value methods SOMEWHAT Employed but NOT fully (in terms of use of forecasts); 3. Earned value methods PARTIALLY employed; 4. Earned value methods MOSTLY employed; 3. Earned value methods FULLY employed			
2.11.5	Labour productivity measurement practice	Refers to the existence of a labour productivity measurement system, separate from Cost control systems.		M	PM
2.11.5.1	Labour productivity process	Refers to whether a system or procedure for measuring daily outputs and work efficiency measures like work sampling is available.	Categorical	I	PM
2.11.5.2	Labour productivity measurement and tracking	Refers to effectiveness of the practice in terms of understanding of labour productivity definition (output to input), use of standard systems to measure quantities (units complete, % complete or level of effort) and work hours, frequency of data collection properly established, productivity evaluation and forecasting (use of performance factor or Earned value method)[32]	1 - 5 Predetermined rating	M	PM
		1. Understanding of labour productivity definition VERY POOR, NO Use of standard systems to measure quantities and work hours, Frequency of data collection NOT properly established, VERY POOR Productivity evaluation and forecasting; 2. Understanding of labour productivity definition POOR, NO Use of standard systems to measure quantities and work hours, Frequency of data collection NOT properly established, POOR Productivity evaluation and forecasting; 3. Understanding of labour productivity definition FAIR, Use of SOME standard systems to measure quantities and work hours, Frequency of data collection SOMEWHAT established, FAIR Productivity evaluation and forecasting; 4. Understanding of labour productivity definition GOOD, Use of SOME standard systems to measure quantities and work hours, Frequency of data collection PROPERLY established, GOOD Productivity evaluation and forecasting; 5. Understanding of labour productivity definition VERY GOOD, Use of PROPER standard systems to measure quantities and work hours, Frequency of data collection PROPERLY established, VERY GOOD Productivity evaluation and forecasting			
2.12.1	Project quality planning	Identifying which quality standards (based on project specifications) are relevant to the project and determining how to satisfy them. Use and understanding of project specification, design requirements, development of clear project quality policy		I	PM
2.12.1	Demand for over quality work	Self-explanatory. Above quality levels set in specifications and drawings.	Categorical (Yes, No)	I	PM

32 Hart, H. (1995).” Measuring construction productivity.” Participant handbook, Construction Industry Institute, Austin, Texas, US.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.12.2	Project quality assurance	Evaluating overall project performance on a regular basis to provide confidence that the project will satisfy the relevant quality standards. Use of quality audits and implementation of quality improvements		I	PM
2.12.1	Quality audits	Refers to the quality audit process in terms of inspections per month.	Real number (Number of inspections per month)	M	PM
2.12.3	Project quality control	Monitoring specific project results to determine if they comply with relevant quality standards and identifying ways to eliminate causes of unsatisfactory performance, as well as identify means to remedy to the non-compliance identified.		I	PM
2.12.3.1	Inspection delay	Refers to the delay caused due to quality control inspections. On average per day.	Real number (Average delay for inspection, min)	D	SI
2.12.3.2	Interference	Interference due to inspections of other trades, safety evaluations, management site visits, measured on average per week. Interference due to inspections of other trades, safety evaluations, management site visits. On average per week.	Real number (Average number of interruption due to interference)	W	SI
2.12.3.3	Inspection programs	Refers to whether the project has a regularly scheduled and coordinated inspection program.	Categorical (Yes, No)	I	SI
2.12.3.4	Out of sequence inspection or survey work	Refers to the occurrence of out sequence inspection or survey works.	Real number (Number of occurrence per week)	W	SI
2.13.1	Procurement planning and solicitation	Refers to the practice of making make-or-buy analysis, developing and selecting alternatives, developing a solicitation and administration plan	1 - 5 Predetermined rating	I	PM
		1. VERY POOR make-or-buy analysis, Developing and selecting alternatives NOT WELL done, VERY POOR solicitation and administration plan; 2. POOR make-or-buy analysis, Developing and selecting alternatives NOT WELL done, POOR solicitation and administration plan; 3. SOME make-or-buy analysis, Developing and selecting alternatives SOMEWHAT done, FAIR solicitation and administration plan; 4. DETAIL make-or-buy analysis, Developing and selecting alternatives WELL done, GOOD solicitation and administration plan; 5. DETAIL make-or-buy analysis, Developing and selecting alternatives WELL done, VERY GOOD solicitation and administration plan			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.13.2	Procurement solicitation planning	Refers to the practice of preparing procurement documents for bids, decisions on contract types, detail evaluation criteria	1 - 5 Predetermined rating	I	PM
		1. VERY POOR procurement documents for bids, VERY POOR Decisions on contract types, LACK of Detail evaluation criteria; 2. POOR procurement documents for bids, POOR Decisions on contract types, LACK of Detail evaluation criteria; 3. FAIR procurement documents for bids, FAIR Decisions on contract types, SOME Detail evaluation criteria; 4. GOOD procurement documents for bids, GOOD Decisions on contract types, Detail evaluation criteria AVALIABLE; 5. VERY GOOD procurement documents for bids, VERY GOOD Decisions on contract types, Detail evaluation criteria AVALIABLE			
2.13.3	Procurement solicitation execution	Refers to the practice on the use of prequalification process, advertisement, evaluation of proposals, award of contract	1 - 5 Predetermined rating	I	PM
		1. NO Use of prequalification process, NO PROPER Advertisement, VERY POOR Practice in evaluation of proposals, NO PROPER Award of contract; 2. NO Use of prequalification process, SOME Advertisement, POOR Practice in evaluation of proposals, NO PROPER Award of contract; 3. SOME Use of prequalification process, SOME Advertisement, FAIR Practice in evaluation of proposals, PROPER Award of contract; 4. DETAIL Use of prequalification process, PROPER Advertisement, GOOD Practice in evaluation of proposals, PROPER Award of contract; 5. DETAIL Use of prequalification process, PROPER Advertisement, VERY GOOD Practice in evaluation of proposals, PROPER Award of contract			
2.13.4	Procurement administration (material, equipment, tool)	Refer to the practice in terms of proper contact with suppliers, placing orders, developing and following deliveries and returns[33]	1 - 5 Predetermined rating	I	PM
		1. INADEQUATE Contact Process, UNORGANIZED Placement of Orders, VERY POOR Follow-up; 2. INADEQUATE Contact Process, UNORGANIZED Placement of Orders, POOR Follow-up; 3. ADEQUATE Contact Process, FAIRLY ORGANIZED Placement of Orders, FAIR Follow-up; 4. ADEQUATE Contact Process, ORGANIZED Placement of Orders, GOOD Follow-up; 5. ADEQUATE Contact Process, WELL ORGANIZED Placement of Orders, VERY GOOD Follow-up			
2.13.5	Trade subcontracting	Refer to the level of subcontracting on the project site Level of Subcontractors on site		I	PM
2.13.5.1	Subcontracted amount	Amount of subcontracted amount.	Real number (%, subcontracted contract amount)	I	PM
2.13.5.2	Number of subcontractors	Total number of subcontractor companies on site.	Real number (Total number of subcontractors per project)	I	PM

[33] Hanna, A. (2012). "Preconstruction planning," *Construction labour productivity management and methods of improvement*, n.p., Madison, Wisconsin, US.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.14.1	Project Safety planning	Development of the approach to manage the various safety hazards inherent to the project. Measured based on understanding of regulatory laws and regulations, contract requirements, detail hazard analysis of project, development of clear project safety plan, development of budget and time for implementation of safety plan, and use of safety officer	1 - 5 Predetermined rating	I	PM
		1. INADEQUATE Understanding of regulatory laws and contract requirements, VERY POORLY done project hazard assessment, INADEQUATE Project Safety plan, VERY POOR Budget and time development; 2. INADEQUATE Understanding of regulatory laws and contract requirements, POORLY done project hazard assessment, INADEQUATE Project Safety plan, POOR Budget and time development; 3. ADEQUATE Understanding of regulatory laws and contract requirements, FAILY done project hazard assessment, ADEQUATE Project Safety plan, FAIR Budget and time development; 4. ADEQUATE Understanding of regulatory laws and contract requirements, ADEQUATELY done project hazard assessment, ADEQUATE Project Safety plan, GOOD Budget and time development; 5. ADEQUATE Understanding of regulatory laws and contract requirements, WELL done project hazard assessment, ADEQUATE Project Safety plan, VERY GOOD Budget and time development			
2.14.1.1	Use of site safety officer	Self-explanatory.	Categorical (Yes, No)	I	PM
2.14.2	Project Safety plan execution	Carrying out the safety plan by performing the activities included in the project safety plan		W	SI
2.14.2.1	Use of daily job hazard assessment forms	Self-explanatory.	Categorical (Yes, No)	I	SI
2.14.2.2	Use of site safety meetings	Use of daily project briefing and debriefing meetings, and tailgate safety meetings	1 - 5 Predetermined rating	M	SI
		1. Safety Meetings NOT conducted; 2. Safety Meetings conducted BUT NOT regularly, Effectiveness of meetings POOR; 3. Safety Meetings conducted REGULARLY, Effectiveness of meetings FAIR; 4. Safety Meetings conducted REGULARLY, Effectiveness of meetings GOOD; 5. Safety Meetings conducted REGULARLY, Effectiveness of meetings VERY GOOD			
2.14.2.3	Construction equipment safety procedure	Availability of proper equipment use procedure	1 - 5 Predetermined rating	M	SI
		1. Proper procedure NOT Available; 2. Proper procedure Available, POOR Implementation of procedure; 3. Proper procedure Available, FAIR Implementation of procedure; 4. Proper procedure Available, GOOD Implementation of procedure; 5. Proper procedure Available, VERY GOOD Implementation of procedure			
2.14.2.4	Drug testing	Self-explanatory.	Categorical	I	SI
2.14.2.5	Safety training	Safety orientation, Fall protection (harness), First aid, H_2S.	Real number (No. trainings attended x Duration of Training, hrs.)	C	F

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.14.2.6	Safety Inspections	Refers to the number of safety inspection per month	Real number (Number of inspections per month)	M	PM
2.14.2.7	Safety Audits	Refers to the number of safety audits per month	Real number (Number of audits per month)	M	PM
2.14.3	Safety Incidents	Refers to the occurrence of safety incidents.		M	SI
2.14.3.1	Near Miss (Unsafe working conditions)	Near Miss - An undesired event that, under slightly different circumstances, could have resulted in personal harm, loss of process, property and/or environmental	Integer (Number of reported near miss per month)	M	SI
2.14.3.2	First Aid (minor personal injury)	First Aid - A first aid is when immediate treatment is rendered by a qualified person and worker immediately returns to work	Integer (Number of reported first aid per month)	M	SI
2.14.3.3	Medical Aid (major personal injury)	Medical Aid - An injury which requires treatment by a physician beyond simple first aid care but does not result in time lost from work beyond the day of the injury	Integer (Number of reported medical aid per month)	M	SI
2.14.3.4	Modified Work Incidents	Modified Work Incident - Work duties which have been modified to accommodate an injured work who cannot perform their regular work duties	Integer (Number of reported modified work incident per month)	M	SI
2.14.3.4	Number of Modified Work Days	Number of Modified Work Days - Days spent performing modified work	Integer (Number of reported modified work days per month)	M	SI
2.14.3.5	Lost Time Incident	Lost Time Incident - Is an accident where a physician directs the injured worker to remain away from work longer that day of the accident	Integer (Number of reported lost time incident per month)	M	SI
2.14.3.6	Number of Lost time Workdays	Number of Lost Time Workdays - Days spent away from work due to accident	Integer (Number of lost day reported due to lost time incident per month)	M	SI
2.14.3.7	Fatality Incident	Self-explanatory.	Integer (Number of reported personnel fatality per month)	M	SI
2.14.3.8	Equipment/Property Damage	Equipment/Property Damage - Accident causing damage to equipment's and/or property on site	Integer (Number of reported equipment/property damage incident per month)	M	SI
2.14.4	Safety Incident investigation	Refers to the practice of carrying out safety investigations.		M	SI
2.14.1	Personnel involved in investigation	Self-explanatory.	Real number (Number of personnel involved in investigation)	M	SI
2.14.2	Process time	Refers to the duration taken in completing the safety investigation.	Real number (Average duration of investigation in hours)	M	SI
2.14.5	Adequacy of Protective gear	Self-explanatory.	Categorical (Yes, No)	I	SI

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.14.6	Uniformity of safety procedures	Self-explanatory.	1 - 5 Predetermined rating	I	SI
		1. VERY Irregular among crews and HIGHLY Variable in daily work times; 2. Irregular among crews and Variable in daily work times; 3. Uniform among crews and Variable in daily work times; 4. Uniform among crews, Always the same in daily work times; 5. VERY Uniform among crews, Always the same in daily work times			
2.14.7	Project Safety administration and reporting	Record keeping of hazard assessment forms, inspections, incidents (near miss, injury, fatality), use of photographs and video records, reporting to project management staff	1 - 5 Predetermined rating	W	SI
		1. VERY POORELY Kept records, NO Use of visual aids; 2. POORELY Kept records, NO Use of visual aids; 3. FAIR Record keeping, Use of visual aids; 4. GOOD Record keeping, GOOD Use of visual aids; 5. VERY GOOD Record keeping, VERY GOOD Use of visual aids			
2.15.1	Risk identification and planning	Refers to proper risk identification, development of an overall risk management plan with risk response planning	1 - 5 Predetermined rating	I	PM
		1. NO Proper risk identification, Development of an overall risk management plan with risk response planning VERY POOR; 2. NO Proper risk identification, Development of an overall risk management plan with risk response planning POOR; 3. SOME Risk identification, Development of an overall risk management plan with risk response planning FAIR; 4. SOME Risk identification, Development of an overall risk management plan with risk response planning GOOD; 5. DETAILED Risk identification, Development of an overall risk management plan with risk response planning VERY GOOD			
2.15.2	Use of risk assessment tool	Refers to the use of qualitative (probability/impact risk rating matrix) or quantitative (Decision tree, simulation, sensitivity analysis) risk assessment tools	1 - 5 Predetermined rating	I	PM
		1. Risk assessment tools NOT used; 2. Risk assessment tools SOMEWHAT used; 3. Risk assessment tools PARTIALLY used; 4. Risk assessment tools MOSTLY used; 3. Risk assessment tools FULLY used			
2.15.3	Risk monitoring and control	Refers to keeping track of identified risks, monitoring residual risks and identifying new risks, ensuring the execution of risk plans, evaluating their effectiveness in reducing risk	1 - 5 Predetermined rating	M	PM
		1. NOT Keeping track of identified risks, VERY POOR Monitoring of residual risks and identifying new risks, VERY POOR in Ensuring the execution of risk plans, NO Evaluation on their effectiveness in reducing risk; 2. NOT Keeping track of identified risks, POOR Monitoring of residual risks and identifying new risks, POOR in Ensuring the execution of risk plans, NO Evaluation on their effectiveness in reducing risk; 3. Keeping SOME track of identified risks, FAIR Monitoring of residual risks and identifying new risks, FAIR in Ensuring the execution of risk plans, SOME Evaluation on their effectiveness in reducing risk; 4. Keeping DETAIL track of identified risks, GOOD Monitoring of residual risks and identifying new risks, GOOD in Ensuring the execution of risk plans, DETAILED Evaluation on their effectiveness in reducing risk; 5. Keeping DETAIL track of identified risks, VERY GOOD Monitoring of residual risks and identifying new risks, VERY GOOD in Ensuring the execution of risk plans, DETAILED Evaluation on their effectiveness in reducing risk			

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.15.4	Crisis management	Refers to understanding possible crises, understanding the time phase of crises (to be reactive or proactive), having systems to prevent crises, understanding stakeholders	1 - 5 Predetermined rating	M	PM
		1. VERY POOR Understanding possible crises and stakeholders, Reactive; 2. POOR Understanding possible crises and stakeholders, Reactive; 3. FAIR Understanding possible crises and stakeholders, Reactive; 4. GOOD Understanding possible crises and stakeholders, Proactive; 5. VERY GOOD Understanding possible crises and stakeholders, Proactive			
2.16.1	Project communication plan	Refers to communication plan, clear roles and responsibilities, identification of stakeholders, distribution of information including reports[34]	1 - 5 Predetermined rating	I	SI
		1. VERY POOR Communication plan, NO Clear roles and responsibilities, NO Identification of stakeholders, VERY POOR Distribution of information; 2. POOR Communication plan, NO Clear roles and responsibilities, NO Identification of stakeholders, POOR Distribution of information; 3. GOOD Communication plan, PROPER Clear roles and responsibilities, PROPER Identification of stakeholders, POOR Distribution of information; 4. GOOD Communication plan, PROPER Clear roles and responsibilities, PROPER Identification of stakeholders, GOOD Distribution of information; 5. VERY GOOD Communication plan, PROPER Clear roles and responsibilities, PROPER Identification of stakeholders, VERY GOOD Distribution of information			
2.16.2	Communication between different trades	Refers to effectiveness of communications between different trades.	1 - 5 Predetermined rating	W	SI
		1. VERY POOR Communication; 2. POOR Communication; 3. FAIR Communication; 4. GOOD Communication; 5. VERY GOOD Communication			
2.16.3	Availability of communication devices	Refers to the number of communication radio devices.	Real number (ratio of communication radio to number of crews, %)	W	SI
2.17.1	Project Interface Development	Refers to the development of site interfaces between project manager, superintendent and foreman with clear project roles and established reporting system	1 - 5 Predetermined rating	I	PM
		1. Interfaces between project team INADEQUATELY developed, NO Clearly established reporting system; 2. Interfaces between project team INADEQUATELY developed, POORELY established reporting system; 3. Interfaces between project team ADEQUATELY developed, FAIRLY established reporting system; 4. Interfaces between project team ADEQUATELY developed, Established reporting system; 5. Interfaces between project team ADEQUATELY developed, WELL Established reporting system			

[34] Awad, A. (2012). “Contractor prequalification using hybrid systems.” PhD thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.17.2	Project staff acquisition	Refers to the recruitment practice of the company		M	SI
2.17.2.1	Hiring practices	Refers to the practice of advertisement, detail job description, reasonable job requirements, fair screening, interview and selection process[35]	1 - 5 Predetermined rating	M	SI
		1. VERY POOR Advertisement, NO Detail job description, Unfair screening, interview and selection process, Unreasonable job requirements; 2. POOR Advertisement, NO Detail job description, Unfair screening, interview and selection process, Unreasonable job requirements; 3. FAIR Advertisement, SOME job description, Unfair screening, interview and selection process, Unreasonable job requirements; 4. GOOD Advertisement, SOME detailed job description, Fair screening, interview and selection process, Reasonable job requirements; 5. VERY GOOD Advertisement, Detailed job description, VERY Fair screening, interview and selection process, Reasonable job requirements			
2.1 7.3	Project team development	Refers to team building activities (picnics, sports contests, holiday outings), reward and recognition systems, trainings		M	PM
2.17.3.1	Team building activities	Use of sport contests, Holiday outings, Picnics, Barbeque events	1 - 5 Predetermined rating	M	PM
		1.Team building events NOT DONE; 2.Team building events DONE, Frequency ATLEAST once per year; 3.Team building events DONE, Frequency ATLEAST twice per year; 4.Team building events DONE, Frequency ATLEAST six times per year; 5.Team building events DONE, Frequency ATLEAST twelve times per year			
2.17.3.2	Reward and recognition system	Reward and recognition for excellence in Safety, Productivity	1 - 5 Predetermined rating	M	SI
		1.Reward and recognition NOT DONE; 2.Reward and recognition DONE, Frequency ATLEAST once per year; 3.Reward and recognition DONE, Frequency ATLEAST twice per year; 4.Reward and recognition DONE, Frequency ATLEAST six times per year; 5.Reward and recognition DONE, Frequency ATLEAST twelve times per year			
2.17.3.3	Work culture	Refers to the work culture in terms of fragmentation, antagonism, mistrust, poor communication, short-term mentality, blame, casual approach to recruitment, machismo and sexism[36]	1 - 5 Predetermined rating	I	PM
[1]	[2]	[3] 1. VERY HIGH Fragmentation, Antagonism, Mistrust, POOR communication, COMMON Short-term mentality, Blame; 2. HIGH Fragmentation, Antagonism, Mistrust, POOR communication, COMMON Short-term mentality, Blame; 3. NORMAL Fragmentation, Antagonism, Mistrust, FAIR communication, COMMON Short-term mentality, Blame; 4. LOW Fragmentation, Antagonism, Mistrust, GOOD communication, UNCOMMON Short-term mentality, Blame; 5. VERY LOW Fragmentation, Antagonism, Mistrust, VERY GOOD communication, UNCOMMON Short-term mentality, Blame			

35 Weiss, D.H. (2004). Fair, square and legal: safe hiring, managing and firing practices to keep you and your company out of court. 4 ed., AMACOM, American Management Association, New York, US.

36 Ankrah, N.A. (2007). “An investigation into the impact of culture on construction project performance.” PhD thesis, University of Wolverhampton, UK.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.17.4	Project team closeout	Layoff practices, Use of personal exit interviews, Development of personnel records		M	PM
2.17.4.1	Use of personal exit interviews	Self-explanatory.	1 - 5 Predetermined rating	I	SI
		1. Exit interview NOT conducted; 2. Exit interview SOMEWHAT conducted; 3. Exit interview PARTIALLY conducted; 4. Exit interview MOSTLY conducted; 5. Exit interview FULLY conducted			
2.17.4.2	Layoff practices	Refers to the layoff practices in terms of reasonable rules, informing the rules to employees, fairness, consistency, follow through[37]	1 - 5 Predetermined rating	M	PM
		1. VERY POOR in informing rules to employees, Unfairness among workers, LACK of Consistency and Follow through, Unreasonable Rules; 2. POOR in informing rules to employees, Unfairness among workers, LACK of Consistency and Follow through, Unreasonable Rules; 3. FAIR in informing rules to employees, Unfairness among workers, GOOD Consistency and Follow through, Reasonable Rules; 4. GOOD in informing rules to employees, Fairness among workers, GOOD Consistency and Follow through, Reasonable Rules; 5. VERY GOOD in informing rules to employees, Fairness among workers, VERY GOOD Consistency and Follow through, Reasonable Rules			
2.18.1	Environmental rating of Project	Refers to environmental rating in terms of LEED (Certified, Silver, Gold, Platinum), BREEAM, BOMA BESt	Categorical (LEED (Certified, Silver, Gold, Platinum), BREEAM, BOMA BESt)	I	PM
2.18.2	Project Environmental Planning	Refers to the practice of understanding of contract environmental provisions, conditions stated in permit applications, project scope statement, project execution characteristics, site and neighborhood condition analysis, development of environmental management plan with Impact analyses and mitigation strategies, use of environmental checklists	1 - 5 Predetermined rating	I	PM
		1. VERY POOR Understanding of contract provisions, Site and neighborhood condition analysis VERY POORLY done, INADEQUATE Environmental management plan, VERY POOR Use of checklists; 2. POOR Understanding of contract provisions, Site and neighborhood condition analysis POORLY done, INADEQUATE Environmental management plan, POOR Use of checklists; 3. FAIR Understanding of contract provisions, Site and neighborhood condition analysis FAILY done, INADEQUATE Environmental management plan, FAIR Use of checklists; 4. GOOD Understanding of contract provisions, Site and neighborhood condition analysis WELL done, ADEQUATE Environmental management plan, GOOD Use of checklists; 5. VERY GOOD Understanding of contract provisions, Site and neighborhood condition analysis VERY WELL done, ADEQUATE Environmental management plan, VERY GOOD Use of checklists			

[37] Hanna, A. (2012). “Preconstruction planning,” *Construction labour productivity management and methods of improvement*, n.p., Madison, Wisconsin, US.

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
2.18.3	Project Environment Assurance	Refers to the practice of sorting waste materials (Concrete, Steel, Wood) and environmental audits		M	PM
2.18.3.1	Environment audits	Self-explanatory.	Real number (No. of inspections per month)	M	PM
2.18.3.2	Sorting of waste materials	Self-explanatory.	1 - 5 Predetermined rating	W	SI
		1. Waste material sorting NOT done; 2. Waste material sorting SOMEWHAT done; 3. Waste material sorting PARTIALLY done; 4. Waste material sorting MOSTLY done; 3. Waste material sorting FULLY done			
2.18.4	Project Environment Control	Refers of the practice of environmental inspections and rework/remedial actions to attain environmental compliance	1 - 5 Predetermined rating	M	PM
		1. NO Use of checklist, NO Rework/remedial action; 2. POOR Use of checklist, NO Rework/remedial action; 3. FAIR Use of checklist, NO Rework/remedial action; 4. GOOD Use of checklist, Rework/remedial action taken when needed; 5. VER GOOD Use of checklist, Rework/remedial action taken when needed			
2.18.5	Rework/Remedial action	Corrective actions taken to meet environmental requirements due to felt impacts like Oil spill	1 - 5 Predetermined rating	M	PM
		1. Corrective action NOT done; 2. Corrective action SOMEWHAT done; 3. Corrective action SOMEWHAT done; 4. Corrective action PARTIALLY done; 5. Corrective action FULLY done			
2.18.6	Environment inspections	Self-explanatory.	Integer (Number of inspections per month)	M	PM
2.18.1	Project claim Identification	Refers to the claim identification with adequacy of claim statements (evidence, contract basis, description of time and cost requirements)	1 - 5 Predetermined rating	M	PM
		1. VERY INADEQUATE; 2. INADEQUATE; 3. FAIRLY ADEQUATE; 4. ADEQUATE; 5. VERY ADEQUATE			
2.18.2	Project claim quantification	Self-explanatory. Experience of claim reviewer, Time take to finalize the review		M	PM
2.18.2.1	Experience of claim reviewer	Self-explanatory.	Real number (Number of years working as claim expert)	M	PM
2.18.2.2	Review process	Self-explanatory.	Real number (Average time taken to finalize a review, weeks)	M	PM
2.18.3	Project claim resolution	Refers to the type of resolution (Negotiation, mediation, arbitration, mini-trials or litigation) and average time taken to resolve the claim		M	PM
2.18.3.1	Resolution method	Refers to the method used, negotiation, mediation, arbitration, mini-trials or litigation. Most frequently used resolution method	Categorical (Negotiation, mediation, arbitration, mini-trials or litigation)	M	PM
2.18.3.2	Resolution process	Average time taken to resolve the claim	Real number (Average time taken to resolve the claim, months)	M	PM

A.4: QUANTIFICATION OF ORGANIZATION LEVEL INPUT PARAMETERS

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
3.1	Organization's principal project type	Defines the types of industries or projects types the organization is seeking	Categorical (Industrial, Commercial, Infrastructure, Institutional, Other)	I	PM
3.2	Organization year in industry	This factor indicates the number of years an organization has been in operation.	Real number (Years in industry)	I	PM
3.3	Annual turnover in dollars	Measure of how much fund turns over year for the organization.	Real number (Annual turnover, million CND$)	I	PM
3.4	Annual employee turnover	Measure of how many employees leave the organization in a year.	Integer (Annual turnover, employee per year)	I	PM
3.5	Organizational structure	Self-explanatory.	Categorical (Matrix, Project based, Mixed)	I	PM
3.6	Project load	Refers to the number of projects handled by the organization in a year.	Integer (Number of projects awarded per year)	I	PM
3.7	Work execution approach	Refers to how much of the work is subcontracted and how much is executed in house.	Real number (%, Ratio of average project amount subcontracted to total project cost)	I	PM

A.5: Quantification of Provincial Level Input Parameters

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
4.1	Provincial economy	Refers to the provincial economy in terms of the annual gross domestic product.	Real number (Provincial GDP, Billion $)	I	DC
4.2	Total number of project	Refers to the number projects similar to the project under study which will compete for resource in the province.	Integer (Number of projects under construction per year in province)	I	DC
4.3	Provincial codes and regulations	Refers to the flexibility of provincial codes and regulations towards the construction industry.	1 - 5 Predetermined rating	I	DC
		1. Most restricted regulations, 2. Strict regulations, 3. Normal regulations, 4. Flexible regulations, 5. Most flexible regulations			
4.4	Unemployment rate	Refers to the annual unemployment rate for construction workers in the province.	Real number (Annual unemployment rate, %)	I	DC
4.5	Labour strikes	Refers to whether a labour force strike related to construction work force was recorded in the year.	Integer (Number of recorded labour strike in construction workforce, annual)	I	DC
4.6	Available supervisor pool in province	Refers to the available workforce qualified for supervision of construction works, at level above a tradesperson and include foreman, superintendent, and project managers.	Integer (Number of qualified supervisors in province, annual)	I	DC
4.7	Tax	Refers to the income tax and goods and services tax levied by the province.		I	DC
4.7.1	Income tax	Refers to the minimum income tax levied by the province.	Real number (minimum income tax, %)	I	DC
4.7.2	GST	Self-explanatory.	Real number (GST, %)	I	DC
4.8	Construction material fluctuation	Refers to the Industrial Product Price Index (IPPI) which measures price changes for major commodities sold by manufacturers in Canada.	Real number (Industrial product price index change, %)	I	DC
4.9	Availability of labour in province	Refers to the total number of trades people in province for activity under study.	Real number (Number of qualified trade workers in the province, annual)	I	DC
4.10	Expenditure level towards projects	Refers to the recorded expenditure towards construction projects in the province.		I	DC
4.10.1	Industrial	Refers to the annual investment made towards industrial projects across the province.	Real number (Annual invested amount, Million $)	I	DC
4.10.2	Commercial	Refers to the annual investment made towards commercial projects across the province.	Real number (Annual invested amount, Million $)	I	DC
4.10.3	Institutional	Refers to the annual investment made towards residential projects across the province.	Real number (Annual invested amount, Million $)	I	DC
4.11	Cost of project (index)	Refers to average cost of building the project per unit of measure like m^2, km, or kW.	Real number (Average cost of project per index)	I	PM

A.6: Quantification of National Level Input Parameters

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
5.1	Political system	Refers to relative stability of the Canadian political system during the initiation and execution of the project.	1 - 5 Predetermined rating	I	PM
		1. VERY unstable; 2. Unstable; 3. Stable; 4. Stable; 5. Very Stable			
5.2	Availability of labour in province	Refers to the total number of trades people in province for activity under study.	Real number (Number of qualified trade workers in the province, annual)	I	DC
5.3	Foreign workers recruitment	Refers to the execution of the foreign workers recruitment program in terms of strictness and processing times.	1 - 5 Predetermined rating	I	DC
		1. VERY STRICT regulations, VERY LONG process time; 2. STRICT regulations, LONG process time; 3. NORMAL regulations, FAIR process time; 4. Flexible regulations, SHORT process time; 5. VERY flexible regulations, VERY SHORT process time			
5.4	Canada population	Refers to the properties of the population in Canada.		I	DC
5.4.1	Size of population	Self-explanatory.	Real number (Population, Million)	I	DC
5.4.2	Growth of population	Self-explanatory.	Real number (Annual growth rate, %)	I	DC
5.4.3	Aging of population	Self-explanatory.	Real number (Median age of Canada's population)	I	DC
5.5	Interest Rates	Refers to average annual interest rate set for prime business by the bank of Canada.	Real number (Annual interest rate, Bank of Canada, %)	I	DC
5.6	Inflation rate	Refers to the average annual inflation rate based on consumer price index (CPI).	Real number (% Change of CPI)	I	DC

A.7: Quantification of Global Level Input Parameters

ID	Parameters	Description	Scale of Measure	Data Cycle	Data Source
6.1	Global economic outlook	Refers to the national economic outlook in terms of real GDP for the coming year, based on the IMF world economy outlook[38]	Real number (Real GDP growth, %)	I	DC
6.2	Global energy supply and demand	Refers to the global energy supply and demand on the previous year.		I	DC
6.2.1	Global energy demand	Self-explanatory.	Real number (Energy demand, Quadrillion BTUs)	I	DC
6.2.2	Global energy supply	Self-explanatory.	Real number (Energy supply, Quadrillion BTUs)	I	DC
6.3	Oil price and price fluctuation	Refers to the current oil price and the weekly fluctuation		D	DC
6.3.1	Oil price	The average WTC (Western Texas Intermediate) oil price is recorded.	Real number (Dollar / barrel)	D	DC
6.3.2	Price fluctuation	The average net price fluctuation on a weekly basis is recorded.	Real number (Weekly price change, %)	W	DC
6.4	Natural gas price and price fluctuation	Refers to the current natural gas and the weekly fluctuation		D	DC
6.4.1	Natural gas price	The average natural gas price is recorded.	Real number (CAD per Million Cubic Feet)	D	DC
6.4.2	Natural gas fluctuation	The average price fluctuation on a weekly basis is recorded.	Real number (Weekly price change, %)	W	DC

38 IMF (2015). “World economic outlook.” International Monetary Fund, < http://www.imf.org/external/pubs/ft/weo/2015/01/ > (June 28, 2015).

Appendix B: Key Input Parameters Influencing CLP

APPENDIX B.1: RANKING AND EVALUATION SCORES OF PARAMETER CATEGORIES POSITIVELY AND NEGATIVELY INFLUENCING CLP

Parameter category	PM survey				Trade survey			
	Building context		Industrial context		Building context		Industrial context	
	Positive influence	Negative influence	Positive influence	Negative influence	Positive influence	Negative influence	Positive influence	Negative influence
A. Labour and crew	9 (38.5)	7 (31.6)	12 (25.1)	6 (45.0)	**3 (92.4)**	7 (12.2)	**1 (100.0)**	6 (20.8)
B. Material and consumables	4 (72.8)	15 (5.4)	**3 (80.8)**	7 (37.2)	7 (43.5)	**1 (100.0)**	9 (28.8)	**2 (93.3)**
C. Equipment and tools	**1 (100.0)**	16 (5.2)	**1 (100.0)**	11 (18.8)	8 (40.8)	**2 (80.4)**	8 (34.5)	**1 (100.0)**
D. Foreman	**2 (96.2)**	17 (2.7)	8 (37.5)	12 (17.3)	**1 (100.0)**	9 (4.4)	**3 (94.9)**	9 (12.8)
E. Task property	14(20.7)	12 (9.8)	15 (6.27)	16 (6.16)	9 (34.5)	5 (49.7)	5 (49.6)	**3 (42.4)**
F. Location property	11 (36.0)	9 (18.3)	7 (40.1)	4 (54.3)	5 (53.5)	4 (57.0)	4 (60.3)	5 (30.8)
G. Project delivery and contract	15 (18.4)	6 (32.6)	16 (3.5)	17 (6.0)	*	*	*	*
H. Engineering and instructions	8 (38.6)	**2 (97.8)**	9 (35.9)	**1 (100.0)**	4 (80.5)	8 (9.6)	7 (45.6)	8 (17.9)
I. Project complexity	17 (5.6)	8 (19.9)	17 (1.31)	15 (8.25)	*	*	*	*
J. Health, safety, and environment	6 (58.5)	10 (17.0)	**2 (87.6)**	14 (12.7)	**2 (92.5)**	**3 (75.7)**	**2 (96.2)**	7 (18.0)
K. Project management practices	5 (63.6)	13 (8.2)	4 (52.9)	9 (22.8)	*	*	*	*
L. Project best practices	7 (55.3)	11 (11.4)	5 (52.8)	10 (22.0)	*	*	*	*
M. Project owner nature	10 (37.6)	5 (43.9)	6 (49.0)	8 (28.3)	*	*	*	*
N. Management of project	*	*	*	*	6 (45.7)	6 (23.2)	6 (48.3)	4 (32.8)
O. Organizational	**3 (75.9)**	14 (6.1)	10 (32.8)	13 (15.6)	*	*	*	*
P. Provincial	13 (25.7)	**3 (61.8)**	13 (23.4)	**3 (88.7)**	*	*	*	*
Q. National	12 (27.0)	4 (57.8	11 (30.6)	5 (47.2)	*	*	*	*
R. Global	16 (6.4)	**1 (100.0)**	14(8.7)	**2 (91.9)**	*	*	*	*

Note: The values in brackets indicate the normalized evaluation score of each category; values in bold represent the top three categories. *Denotes that the parameter category is not included in the survey

APPENDIX B.2: PERSPECTIVE ANALYSIS ON PM AND TRADE RESPONDENT GROUPS WITHIN THE SAME CONTEXT

Building context Parameter	Evaluation score		Δ	F-value	Industrial context Parameter	Evaluation score		Δ	F-value
	PM	Trade				PM	Trade		
Positive influence:					**Positive influence:**				
Crew is experienced and has the necessary competence	5.3	70.1	64.8	14.9[a]	There is a really good cooperation between craftsmen in a crew	2.5	100.0	97.5	4.4[a]
Frequency of accidents and personal injury is low	31.3	88.8	57.5	24.9[a]	Crew is experienced and has the necessary competence	4.1	62.0	57.8	19.7[a]
Work is fairly assigned to the different crews	26.3	67.3	41.1	1.8	Crew is given adequate training before commencement	4.8	48.9	44.1	0.2
Drawings and specifications are readily available	22.6	60.5	38.0	1.3	In this project, rework is not frequent	1.4	39.6	38.3	16.2[a]
Workers can get the required hand tools to do their jobs	56.2	19.7	36.6	18.0[a]	Craftsmen are properly treated by foreman	2.9	39.1	36.2	3.2
Negative influence:					**Negative influence:**				
Stringent safety rules are negatively affecting productivity	4.5	74.9	70.4	2.7	Materials are not delivered on time to task location	10.1	96.1	86.0	1.7
I wait in a line for manlifts	1.3	30.6	29.4	0.2	Workers can't get the required hand tools to do their jobs	0.2	79.1	78.9	47.5[a]
On average the weather is harsh (temperature, wind, humidity, precipitation)	7.9	33.9	26.0	12.7[a]	The materials delivered have quality problems	0.6	69.6	69.1	17.0[a]
In this project, interruption and disruption are frequent	2.8	21.4	18.7	0.0	There is a shortage of good transportation equipment (cranes, forklifts)	1.8	46.0	44.2	6.2[a]
There is frequent crew turnover	5.9	23.7	17.8	2.3	The site does not have a very good material order tracking system	0.8	40.0	39.2	19.1[a]

Note: Sample sizes for the PM survey in the building and industrial contexts were 20 and 22, respectively, and for the trade survey in the building and industrial contexts were 58 and 41, respectively. [a] Indicates the difference between the PM and trade respondent groups is significant at the 95% confidence level.

APPENDIX B.3: PERSPECTIVE ANALYSIS OF BUILDING AND INDUSTRIAL CONTEXTS USING PM AND TRADE RESPONDENT GROUPS

PM survey respondents Parameter	Evaluation score Building	Evaluation score Industrial	Δ	F-value	Trade survey respondents Parameter	Evaluation score Building	Evaluation score Industrial	Δ	F-value
Positive influence:					**Positive influence:**				
There are adequate and quality work tools	100.0	12.5	87.5	5.6 [a]	The work area is protected from weather effect	5.1	60.7	55.6	33.0[a]
Daily job hazard assessment system is in place	82.6	4.2	78.4	1.2	Craftsmen’s labour union status (unionized or not unionized) and its benefits are important in their day to day performance	13.3	67.9	54.6	4.5 [a]
Project site safety rules are not stringent	4.2	68.3	64.2	1.1	Washrooms are closely located	55.8	9.4	46.3	51.4 [a]
Efforts are taken to reduce turnover of foremen	72.6	10.0	62.6	8.0 [a]	Work permits are provided in a timely fashion	7.6	51.6	44.0	3.8
Integration management practices: The process of coordinating the various elements of the project is properly implemented	69.5	10.0	59.5	10.4 [a]	Work is fairly assigned to the different crews	67.3	23.6	43.8	35.2 [a]
Negative influence:					**Negative influence:**				
There are many competing projects within the province	43.9	100.0	56.1	9.9 [a]	Stringent safety rules are negatively affecting productivity	74.9	18.6	56.4	0.4
Drawings and specifications are often not complete and require updates	54.4	2.1	52.3	0.1	Workers cannot get the required power tools to do their jobs	34.7	90.8	56.1	25.0 [a]
Crew experience and competence is not meeting expectations	42.0	2.2	39.8	3.6	Workers cannot get the required hand tools to do their jobs	33.8	79.1	45.3	18.4 [a]
Prices for outputs (project completion costs) are substantially increasing	37.4	76.7	39.3	1.3	The site does not have a very good material order tracking system	1.1	43.8	42.8	46.6 [a]
Drawings and specifications are not made available well ahead of implementations	1.4	40.4	39.0	5.1	Work conditions are compromised by excessive noise, dust and fumes	1.6	40.0	38.4	23.8 [a]

Note: Sample sizes for the PM survey in the building and industrial contexts were 20 and 22, respectively, and for the trade survey in the building and industrial contexts were 58 and 41, respectively. [a] Indicates the difference between the building and industrial PM and trade respondent groups is significant at the 95% confidence level

Appendix B.4: Pearson Correlation: Key Parameters Influencing CLP Using Data-driven Approach

	x11	x12	x13	x15	x23	x37	x58	x96	x97	x101	x126	x143	x45	x46	z = (CLP)
x11	1														
x12	-0.050	1													
	(0.633)														
x13	-0.491 [a]	-0.262 [a]	1												
	(0.000)	(0.011)													
x15	-0.004	0.064	0.324 [a]	1											
	(0.968)	0.547	0.002												
x23	0.179	0.010	-0.174	-0.140	1										
	(0.088)	(0.926)	(0.098)	(0.182)											
x37	-0.171	0.011	0.050	-0.019	0.001	1									
	(0.104)	(0.914)	(0.637)	(0.856)	(0.990)										
x58	0.094	-0.064	-0.149	-0.034	-0.140	-0.019	1								
	(0.373)	(0.545)	(0.156)	(0.750)	(0.182)	(0.856)									
x96	-0.775 [a]	-0.031	0.515 [a]	-0.087	0.006	0.220 [a]	-0.087	1							
	(0.000)	(0.768)	(0.000)	(0.407)	(0.955)	(0.035)	(0.407)								
x97	0.436 [a]	-0.090	-0.476 [a]	-0.121	-0.079	-0.069	0.278 [a]	-0.315 [a]	1						
	(0.000)	(0.393)	(0.000)	(0.249)	(0.452)	(0.511)	(0.007)	(0.002)							
x101	-0.074	-0.030	-0.042	-0.016	-0.243 [a]	-0.009	0.345 [a]	-0.041	0.501 [a]	1					
	(0.481)	(0.778)	(0.692)	(0.882)	(0.020)	(0.932)	(0.001)	(0.700)	(0.000)						
x126	0.091	-0.323 [a]	0.210 [a]	0.065	-0.132	-0.061	-0.106	-0.275 [a]	-0.382 [a]	-0.209 [a]	1				
	(0.388)	(0.002)	(0.045)	(0.537)	(0.209)	(0.5660	(0.314)	(0.008)	(0.000)	(0.046)					
x143	0.417 [a]	-0.198	-0.468 [a]	-0.062	0.008	-0.091	0.230 [a]	-0.413 [a]	0.842 [a]	0.416 [a]	-.0207 [a]	1			
	(0.000)	(0.058)	(0.000)	(0.558)	(0.943)	(0.388)	(0.027)	(0.000)	(0.000)	(0.000)	(0.047)				
x45	-0.295 [a]	0.471 [a]	0.116	-0.115	-0.086	0.129	-0.115	0.431 [a]	-0.151	0.026	-0.363 [a]	-0.287 [a]	1		
	(0.004)	(0.000)	(0.270)	(0.274)	(0.413)	(0.219)	(0.274)	(0.000)	(0.151)	(0.806)	(0.000)	(0.006)			
x46	-0.014	0.078	-0.174	-0.132	-0.066	0.101	0.083	0.005	0.086	-0.004	0.007	0.065	0.071	1	
	(0.892)	(0.462)	(0.097)	(0.208)	(0.531)	(0.340)	(0.429)	(0.964)	(0.413)	(0.968)	(0.945)	(0.540)	(0.501)		
z (CLP)	-0.261 [a]	-0.143	0.273 [a]	0.099	-0.073	0.091	-0.135	0.227 [a]	-0.343 [a]	-0.175	0.218 [a]	-0.289 [a]	-0.078	0.227 [a]	1
	(0.012)	(0.173)	(0.008)	(0.346)	(0.487)	(0.387)	(0.199)	(0.030)	(0.001)	(0.095)	(0.037)	(0.005)	(0.461)	(0.030)	

Note: Values in parentheses indicate the significant value for a two-tailed correlation test. [*] Denotes a statistically significant correlation at a significance level of 0.05.

Appendix C: Work Sampling Proportion versus CLP

APPENDIX C.1: CLP AS FUNCTION OF DIRECT WORK PROPORTION: SCATTER PLOTS

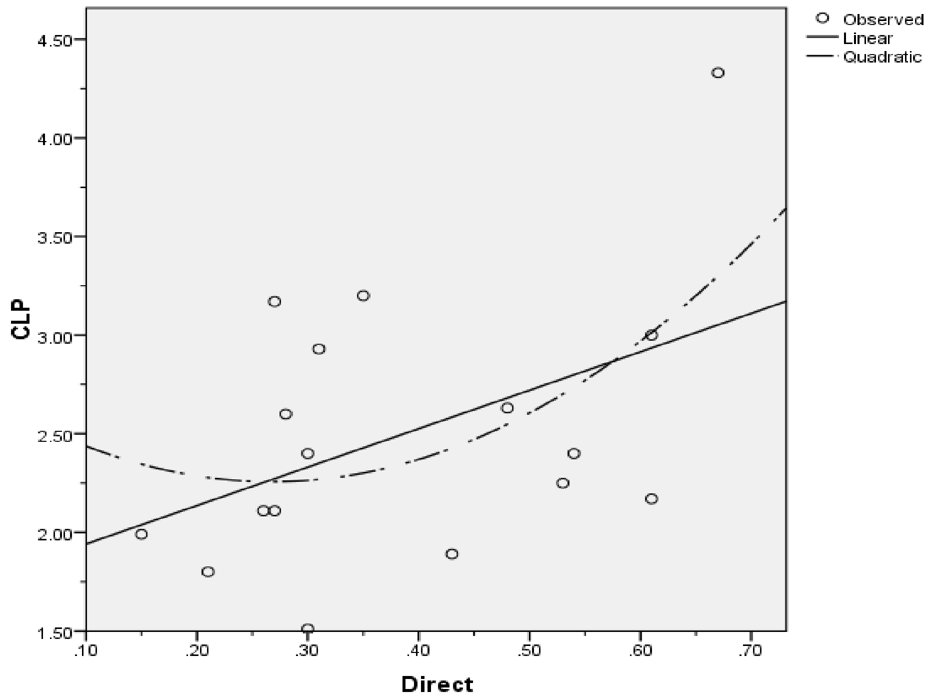

Figure C1.1: CLP as Function of Direct Work Proportion for Column Concreting Activity

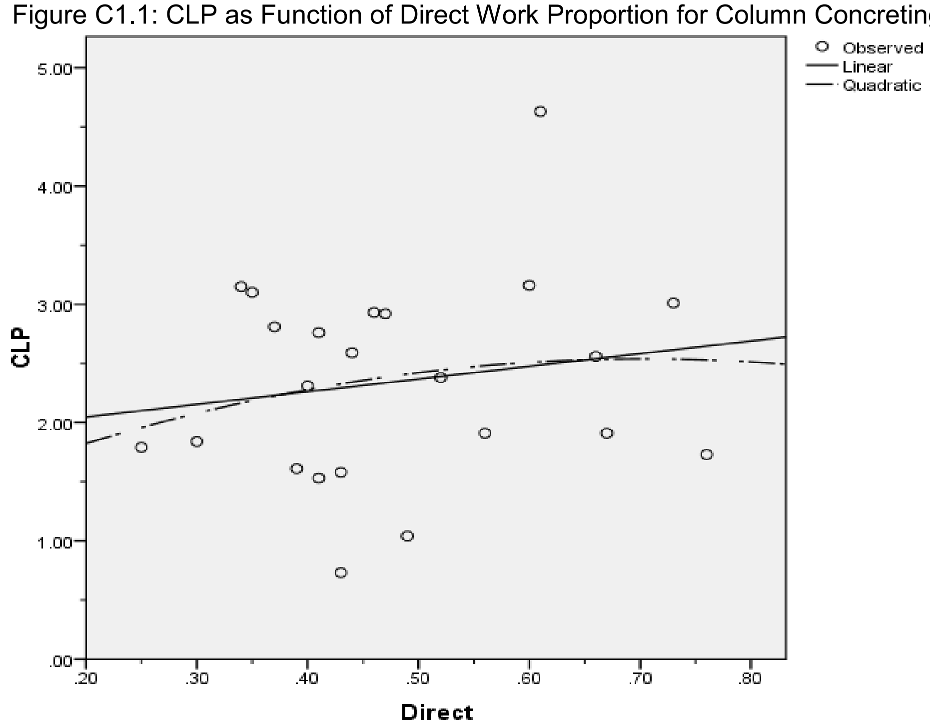

Figure C1.2: CLP as Function of Direct Work Proportion for Slab Concreting Activity

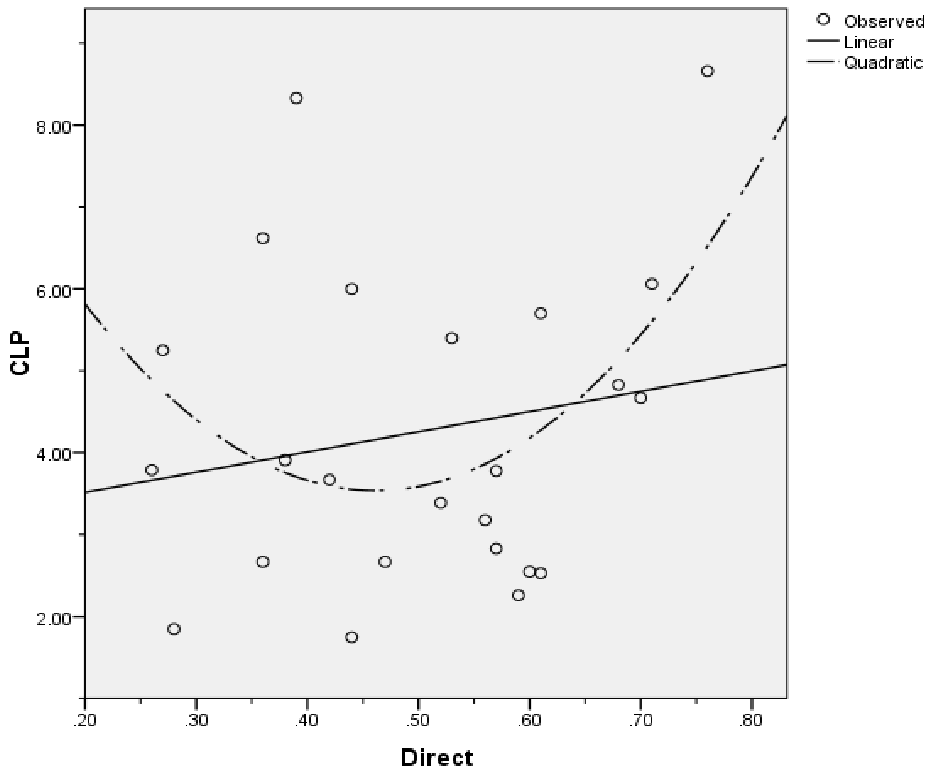

Figure C1.3: CLP as Function of Direct Work Proportion for Wall Concreting Activity

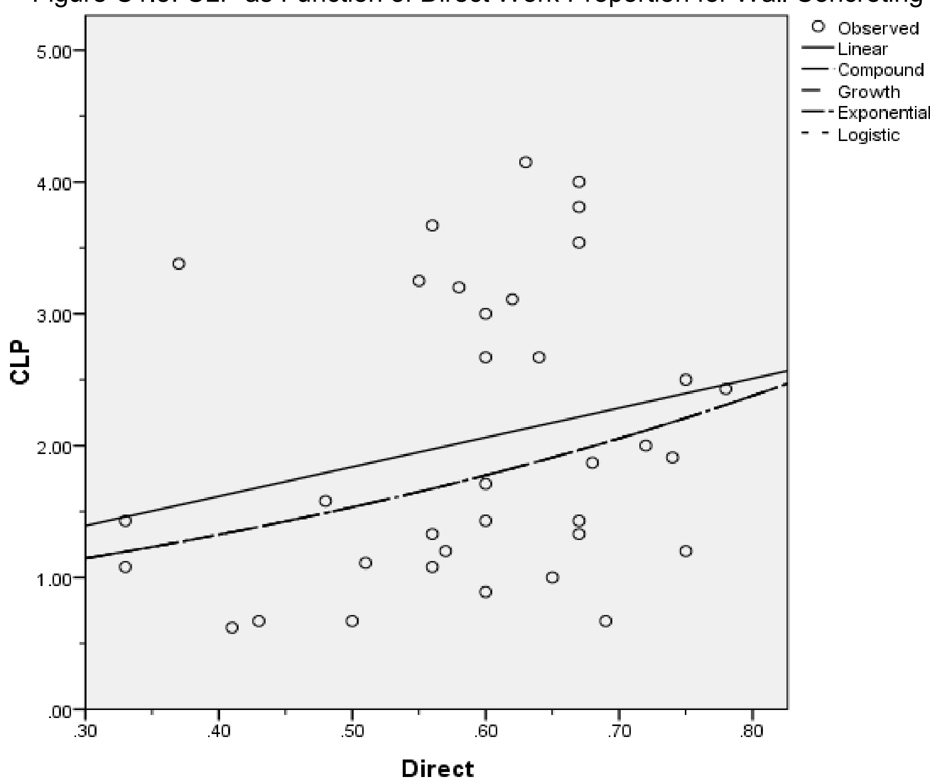

Figure C1.4: CLP as Function of Direct Work Proportion for Box Installation Activity

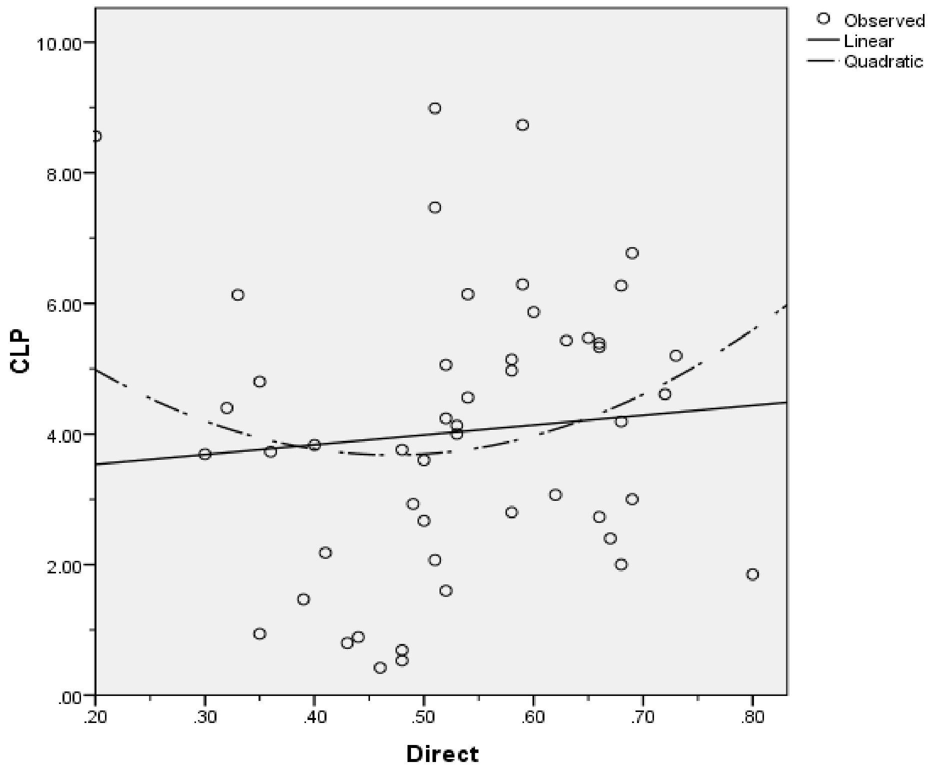

Figure C1.5: CLP as Function of Direct Work Proportion for Piping Activity

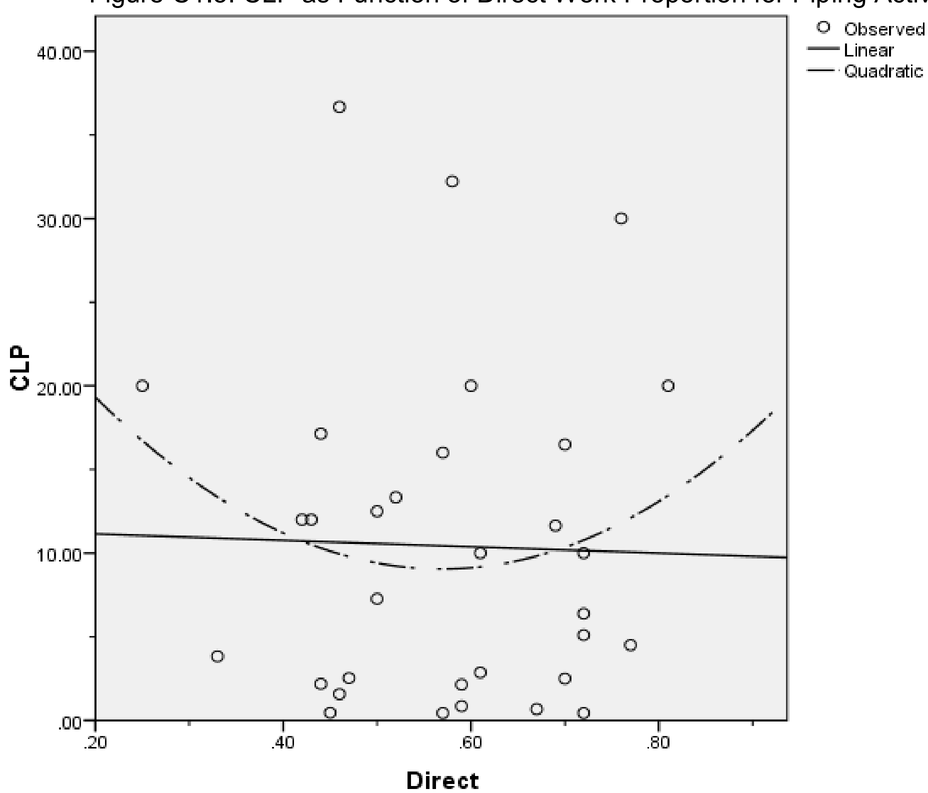

Figure C1.6: CLP as Function of Direct Work Proportion for Wire Pulling Activity

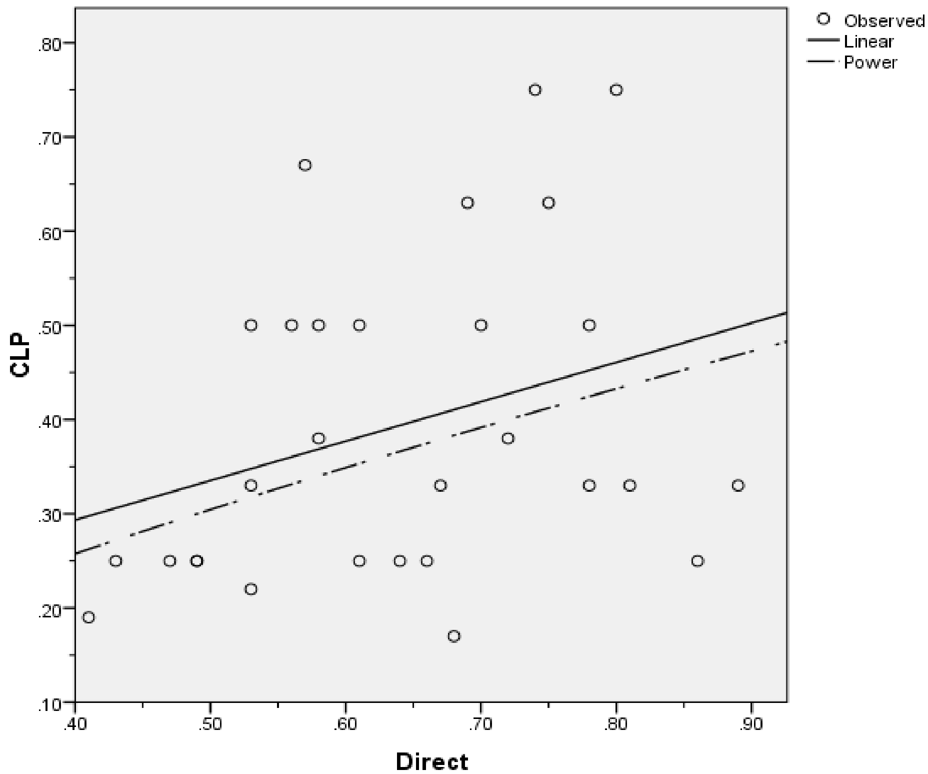

Figure C1.7: CLP as Function of Direct Work Proportion for Overlay Activity

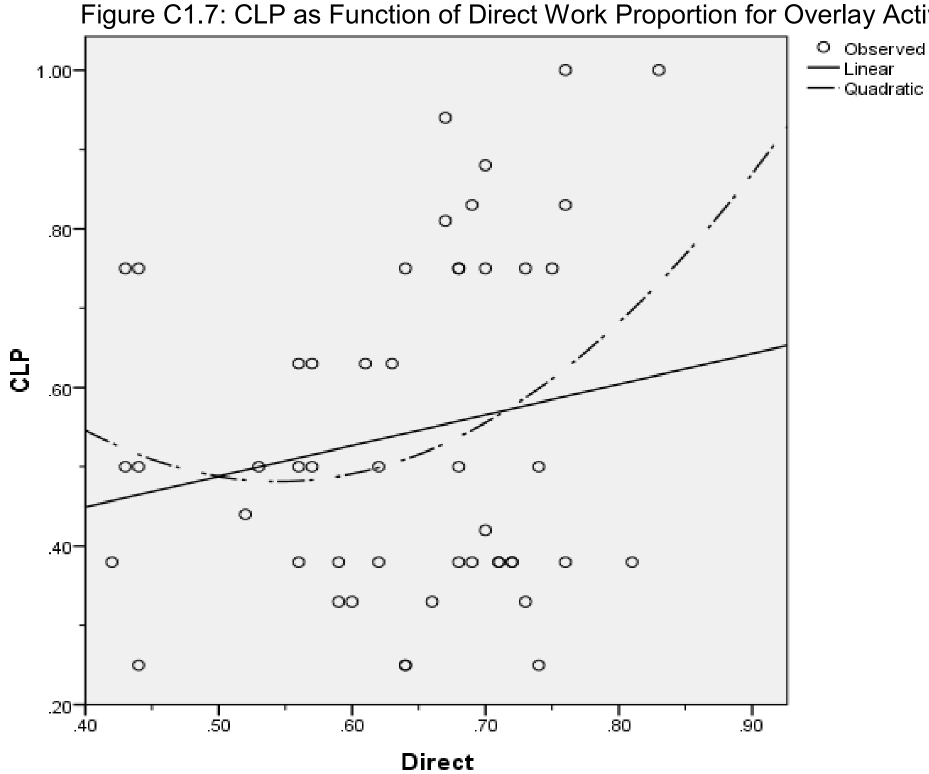

Figure C1.8: CLP as Function of Direct Work Proportion for Shields Installation Activity

APPENDIX C.2: PEARSON CORRELATION ANALYSIS: WORK SAMPLING PROPORTIONS WITH CLP

Table C1.1: Pearson Correlation: Work Sampling Proportion (%) with CLP for Column Concreting Activity

WS and CLP	Correlation coefficient of independent variables							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.206 (0.427)	1						
y_3	0.392 (0.120)	-0.270 (0.295)	1					
y_4	-0.700 [a] (0.002)	0.241 (0.352)	−0.490 [a] (0.046)	1				
y_5	-0.618 [a] (0.008)	-0.492 [a] (0.045)	-0.409 (0.103)	0.238 (0.358)	1			
y_6	-0.552 [a] (0.022)	-0.298 (0.245)	-0.482 (0.050)	0.281 (0.358)	0.714 [a] (0.001)	1		
y_7	c (c)	c (c)	c (c)	c (c)	c (c)	c (c)	1	
Z	0.454 (0.067)	-0.398 (0.114)	-0.118 (0.651)	-0.109 (0.677)	-0.035 (0.895)	0.043 (0.871)	c (c)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test and *c* indicates that the value cannot be computed as the data instance values are constant.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.2: Pearson Correlation: Work Sampling Proportion (%) with CLP for Slab Concreting Activity

WS and CLP	Correlation coefficient of independent variables							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	0.346 (0.124)	1						
y_3	-0.362 (0.107)	-0.350 (0.120)	1					
y_4	0.024 (0.917)	-0.148 (0.521)	0.104 (0.654)	1				
y_5	-0.873 [a] (0.000)	-0.275 (0.228)	0.137 (0.555)	-0.159 (0.491)	1			
y_6	0.592 [a] (0.005)	0.306 (0.177)	-0.124 (0.593)	0.056 (0.809)	-0.674 [a] (0.001)	1		
y_7	-0.425 (0.055)	-0.340 (0.131)	-0.175 (0.448)	0.137 (0.555)	0.108 (0.641)	-0.205 (0.373)	1	
Z	0.104 (0.652)	0.278 (0.222)	-0.169 (0.464)	0.398 (0.074)	-0.057 (0.806)	0.093 (0.689)	-0.116 (0.618)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.3: Pearson Correlation: Work Sampling Proportion (%) with CLP for Wall Concreting Activity

WS and CLP	**Correlation coefficient of independent variables**							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.133 (0.535)	1						
y_3	-0.490 [a] (0.015)	-0.227 (0.287)	1					
y_4	-0.198 (0.354)	-0.013 (0.953)	-0.391 (0.059)	1				
y_5	-0.625 [a] (0.001)	-0.230 (0.280)	-0.098 (0.647)	0.229 (0.281)	1			
y_6	-0.082 [a] (0.704)	0.269 (0.204)	-0.089 (0.680)	-0.143 (0.504)	0.006 (0.979)	1		
y_7	c (c)	c (c)	c (c)	c (c)	c (c)	c (c)	1	
Z	0.186 (0.384)	0.001 (0.997)	0.003 (0.989)	-0.465 (0.022)	-0.041 (0.850)	0.024 (0.912)	c (c)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test and *c* indicates that the value cannot be computed as the data instance values are constant.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.4: Pearson Correlation: Work Sampling Proportion (%) with CLP for Box Installation Activity

WS and CLP	**Correlation coefficient of independent variables**							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.484 [a] (0.004)	1						
y_3	-0.208 (0.237)	-0.079 (0.659)	1					
y_4	-0.313 (0.072)	-0.228 (0.194)	0.296 (0.090)	1				
y_5	0.000 (0.998)	0.125 (0.481)	-0.171 (0.333)	0.064 (0.721)	1			
y_6	-0.283 (0.105)	-0.062 (0.728)	0.258 (0.141)	0.188 (0.288)	0.012 (0.001)	1		
y_7	-0.425 (0.055)	-0.340 (0.131)	-0.175 (0.448)	0.137 (0.555)	0.108 (0.641)	-0.205 (0.946)	1	
Z	0.217 (0.219)	-0.343 (0.047)	-0.095 (0.592)	0.165 (0.351)	0.149 (0.399)	-0.158 (0.373)	0.018 (0.918)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.5: Pearson Correlation: Work Sampling Proportion (%) with CLP for Piping Activity

WS and CLP	Correlation coefficient of independent variables							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.537 (0.000)	1						
y_3	-0.077 (0.595)	-0.380 (0.006)	1					
y_4	-0.409 (0.003)	0.207 (0.149)	-0.168 (0.244)	1				
y_5	-0.233 (0.104)	0.273 (0.055)	0.102 (0.482)	-0.135 (0.349)	1			
y_6	-0.441 [a] (0.001)	0.028 (0.849)	-0.061 (0.673)	0.143 (0.323)	0.045 (0.754)	1		
y_7	-0.508 (0.000)	0.048 (0.743)	-0.136 (0.347)	-0.089 (0.541)	-0.050 (0.728)	0.138 (0.340)	1	
Z	0.090 (0.535)	-0.330 (0.019)	0.208 (0.147)	-0.119 (0.412)	-0.201 (0.161)	-0.025 (0.864)	0.030 (0.836)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.6: Pearson Correlation: Work Sampling Proportion (%) with CLP for Wire Pulling Activity

WS and CLP	Correlation coefficient of independent variables							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.324 (0.071)	1						
y_3	-0.182 (0.319)	0.351 (0.049)	1					
y_4	-0.252 (0.163)	0.465 [a] (0.007)	0.310 (0.084)	1				
y_5	-0.041 (0.826)	0.345 (0.053)	0.176 (0.335)	0.238 (0.190)	1			
y_6	0.001 (0.994)	0.264 (0.145)	0.406 [a] (0.021)	0.102 (0.580)	0.147 (0.421)	1		
y_7	0.577 [a] (0.001)	-0.486 [a] (0.005)	-0.339 (0.057)	-0.368 (0.038)	-0.224 (0.219)	-0.461 (0.008)	1	
Z	-0.027 (0.884)	-0.027 (0.882)	-0.098 (0.594)	-0.125 (0.497)	-0.214 (0.240)	-0.076 (0.680)	0.089 (0.628)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.7: Pearson Correlation: Work Sampling Proportion (%) with CLP for Overlays Activity

WS and CLP	Correlation coefficient of independent variables							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.789 [a] (0.000)	1						
y_3	-0.189 (0.325)	0.124 (0.523)	1					
y_4	-0.374 [a] (0.046)	0.159 (0.409)	0.053 (0.786)	1				
y_5	-0.333 (0.077)	0.126 (0.515)	-0.470 (0.010)	0.220 (0.252)	1			
y_6	-0.296 (0.118)	-0.032 (0.870)	-0.187 (0.331)	0.135 (0.487)	-0.067 (0.731)	1		
y_7	-0.115 (0.553)	-0.381 [a] (0.042)	-0.135 (0.487)	0.086 (0.656)	-0.067 (0.731)	0.568 (0.001)	1	
Z	0.317 (0.094)	-0.209 (0.276)	-0.054 (0.779)	-0.272 (0.153)	0.114 (0.556)	-0.334 (0.076)	-0.251 (0.189)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test.
[a] Denotes a statistically significant correlation at a significance level of 0.05.

Table C1.8: Pearson Correlation: Work Sampling Proportion (%) with CLP for Shields Installation Activity

WS and CLP	Correlation coefficient of independent variables							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Z
y_1	1							
y_2	-0.565 (0.000)	1						
y_3	-0.214 (0.139)	-0.133 (0.361)	1					
y_4	0.024 (0.872)	-0.374 (0.008)	0.049 (0.737)	1				
y_5	-0.607 [a] (0.000)	0.021 (0.888)	-0.201 (0.166)	0.087 (0.553)	1			
y_6	-0.460 [a] (0.001)	0.223 (0.124)	0.081 (0.580)	-0.079 (0.590)	0.061 (0.679)	1		
y_7	-0.405 [a] (0.004)	0.134 (0.360)	-0.165 (0.257)	-0.147 (0.313)	0.326 [a] (0.022)	-0.087 (0.554)	1	
Z	0.190 (0.192)	0.005 (0.971)	-0.270 (0.060)	0.077 (0.600)	-0.038 (0.796)	-0.087 (0.551)	-0.123 (0.399)	1

Note: Values in parentheses indicate the significant value for a two-tailed correlation test.
[a] Denotes a statistically significant correlation at a significance level of 0.05.