Data-Driven Strategies to Improve the Construction Equipment Management

by

Chang Liu

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 equipment management is critical for the long-term success of construction companies. Managing equipment in a cost-efficient manner for project or corporate operations is a key concern for construction companies. Although equipment cost is normally simplified to be a unit rate in project bidding and management, it is actually an aggregate of numerous small components. To maintain a competitive edge, construction companies need to analyze these small components and obtain cost- or time-saving strategies to enhance their management performance and decision making. The goal of this research is to introduce a new generation of data-driven, simulation-based analytics for construction equipment management to provide analytical decision support to industrial practitioners.

Current construction equipment management requires both experience and expertise. Data plays a vital role in assisting decision making for equipment management. Vast amounts of data are available today, especially for the equipment costs and location-tracking, but only a small portion has been used. Additionally, simplified analytical tools used in some management strategies overlook valuable information and enhance data collected through processing, structuring, and interpretation. To address the limitations of current practices, this research created data-driven, simulation-based analytics to provide decision support to construction equipment management as follows: 1) dynamic quantification methods to achieve bargains in equipment trading; 2) simulation-based life-cycle cost analysis for heavy equipment; and 3) performance measurement method for equipment logistics.

For the input modeling, K-means clustering and the Expectation-Maximization (EM) algorithm were used to obtain the distributions of inputs. To achieve dynamic updating, Bayesian inference was applied, integrating newly-generated and historical data to re-calibrate the inputs. Markov Chain Monte Carlo (MCMC) method was employed to approximate the posterior distribution after Bayesian inference. For the analytics, mathematical modelling was applied, and social network analysis (SNA) was introduced to evaluate equipment dispatch. Life-cycle cost analysis (LCCA) was also applied to incorporate both maintenance and ownership costs. Feasibility and functionality of the proposed research was validated through practical case studies. These case studies demonstrated the applications of proposed simulation-based analytics in detail and provided valuable information for practitioners. These approaches have been shown to be effective in achieving bargains in equipment acquisition and disposal, predicting the cumulative total cost of equipment, and evaluating equipment logistics performance, all of which can provide analytical decision support for equipment-management practitioners.

PREFACE

This thesis is the original work of Chang Liu. It is organized in paper format and is written on the basis of the related research papers listed below:

1. A version of Chapter 2 has been published as Liu, C., Lei, Z., Morley, D. and AbouRizk, S. (2020). "*Dynamic, Data-Driven Decision-Support Approach for Construction Equipment Acquisition and Disposal*." Journal of Computing in Civil Engineering, *34*(2). Dr. Lei was involved in concept formation and manuscript composition. Mr. Morley was involved in concept formation. Dr. AbouRizk was the supervisory authority and was involved in manuscript composition.

2. A version of Chapter 3 has been published as Liu, C., AbouRizk, S., Morley, D. and Lei Z. (2020). "*Data-Driven Simulation-Based Analytics for Heavy Equipment Life-Cycle Cost.*" Journal of Construction Engineering and Management. 146(5). Dr. AbouRizk was the supervisory authority and was involved in manuscript composition. Mr. Morley was involved in concept formation. Dr. Lei was involved in concept formation and manuscript composition.

3. A version of Chapter 4 has been published as Liu, C., Ji, W., AbouRizk, S. and Siu, M.F.F. (2019). "*Equipment Logistics Performance Measurement Using Data-Driven Social Network Analysis*." Journal of Construction Engineering and Management, 145(5). Dr. Ji and Dr. Siu were involved in concept formation and manuscript composition. Dr. AbouRizk was the supervisory authority and was involved in manuscript composition.

ACKNOWLEDGEMENTS

To my supervisor: Dr. Simaan M. AbouRizk.

To my committee members: Dr. Ahmed Hammad, Dr. Ali Imanpour, Dr. Mohamed Al-Hussein, Dr. Ming Lu, Dr. Ian Buchanan, and Dr. Samuel Ariaratnam.

To my friends: Dr. Ming Fung Siu, Dr. Wenying Ji, and Dr. Zhen Lei

To my colleagues: Dr. Catherine Pretzlaw, Dr. Mickey Richards, Mr. Stephen Hague, and Ms. Brenda Penner.

To my family and beloved wife: Ms. Yihan Zhao.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	**Meaning**
ANN	Artificial neural network
ART	Autoregressive tree algorithm
ASAE	American Society of Agricultural Engineers
B	Bargain percentage
BIM	Building information modeling
BU	Business unit(s)
CAT	Caterpillar
CCM	Cumulative cost model
CEM	Construction engineering and management
DDDAS	Dynamic data-driven application system
DDI	Direct dispatch index
DES	Discrete event simulation
EM	Expectation-Maximization algorithm
ERP	Enterprise resource planning
GPS	Global positing system
IQR	Interquartile range
KNN	K-nearest neighbor
LCCA	Life-cycle cost analysis
LTD	Life-to-date
MC	Maintenance cost
MCMC	Markov chain Monte Carlo
MLE	Maximum likelihood estimate
OC	Ownership cost
PCB	Period-cost-based
PPP	Public-private-partnership

Q	Quantity of equipment transaction
RF	Random forest
RMV	Residual market value
RSE	Relative square error
RRSE	Root relative squared error
SMR	Service meter reading
SNA	Social network analysis
TC	Total cost

1 CHAPTER 1: INTRODUCTION

1.1 Research Background

Equipment is a major resource that is vital to the success of every construction project. The successful completion of construction projects depends on reliable and functional equipment fleets (Vorster 2009). In Canada, equipment costs can account for up to 60% of the total cost of many construction projects (University of Toronto 2001). At the project level, equipment needs to be efficiently managed at the jobsite in a timely and cost-effective manner. To achieve long-term success at the corporate level, equipment needs to be well-maintained, and as a financial asset, it should be acquired or disposed of at the right time, which can be challenging. In general, construction equipment management requires inter-disciplinary knowledge, involving construction engineering and management (CEM), equipment economics and mechanical engineering. For example, selection and operation of mobile cranes in industrial construction typically requires both heavy lift studies from construction engineers and the equipment's mechanical parameters from equipment managers (Lei 2015). Construction equipment management relies heavily on quantifications and analyses to make decisions in all aspects of equipment management, including equipment operation, utilization and costs.

Construction equipment management, for general contractors or municipal governments, includes operational, financial and mechanical decisions for the equipment. Cost-efficient equipment management can facilitate construction operations and enhance profits at the corporate level. To achieve long-term profits, the structure of the equipment management department or division is typically determined based on corporate size, among other factors. Two major structures of equipment management in corporations are decentralized equipment management and centralized equipment management.

Decentralized equipment management means that the corporation does not have any equipment management business unit (BU) to oversee equipment management. The project management teams (operational BUs) are responsible for equipment maintenance and financial reporting of equipment costs. When the number of projects managed by the corporation increases, decentralized equipment management becomes more complex. It requires that each project management team has an insightful understanding of equipment management, which may not be easily achieved and managed at the corporate level, especially for large general contractors.

Centralization of equipment management is becoming more popular, especially for large projects over multiple provinces or states. A centralized equipment management team, instead of multiple project management teams, is responsible for all equipment costs related to ownership, maintenance and equipment use. Whether construction equipment is generating profits can be easily examined by the centralized equipment management team and executive managers. In some cases, equipment revenues by different disciplines (e.g. heavy civil, building and industrial) are further assessed, which is difficult in decentralized equipment management but achievable through centralized equipment management. One typical structure of centralized equipment management structure is illustrated in the Figure 1.1.

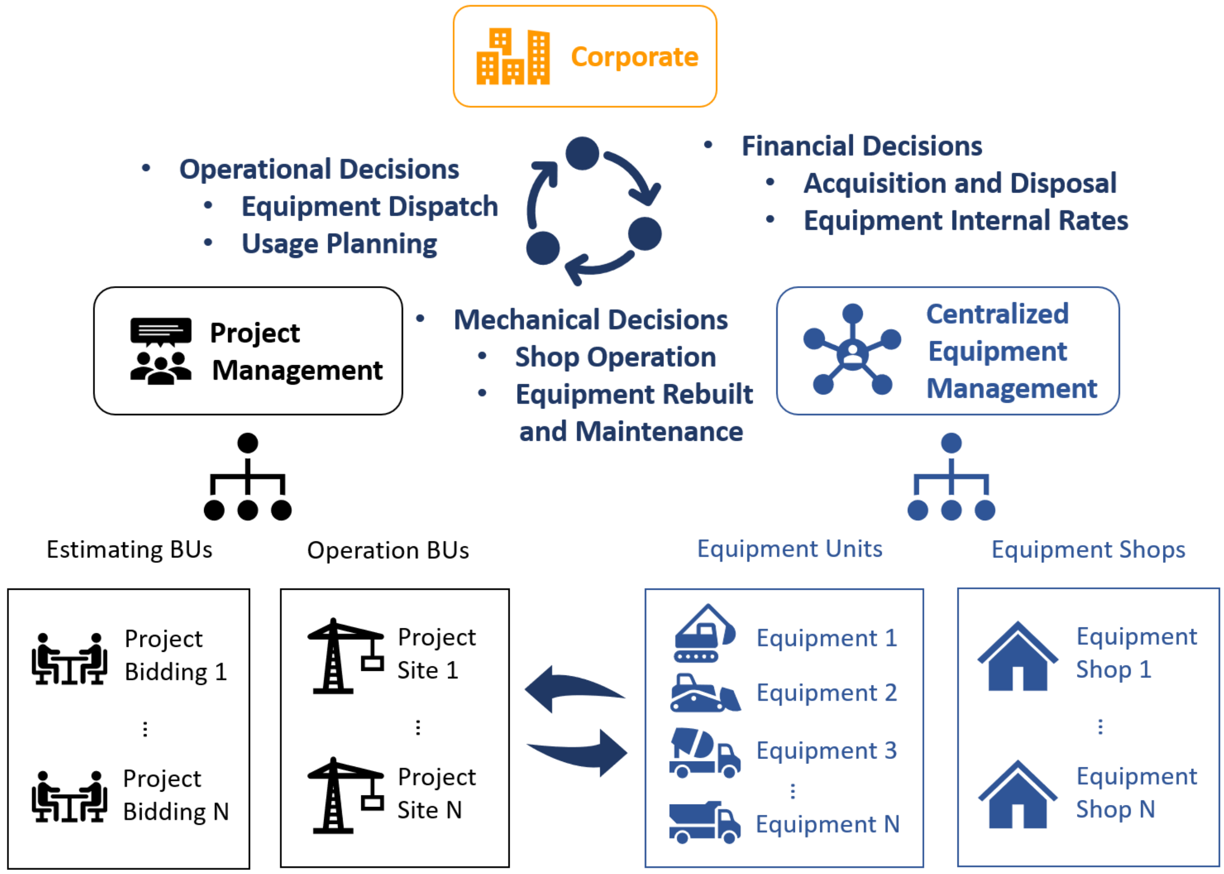


Figure 1.1 A Typical Centralized Equipment Management Structure

Centralized equipment management, as shown in Figure 1.1, makes 1) financial decisions, including equipment acquisition/disposal, and sets internal costs for equipment (e.g. dollars per hour or dollars per day); 2) mechanical decisions, including rebuilding equipment, preventative and routine maintenance, and shop operations; and 3) operational decisions for equipment dispatch and equipment usage planning.

To properly manage construction equipment, decision-makers require quantification and analyses of historical and dynamic data. At the corporate level, equipment cost is most frequently examined to ensure financial sustainability. However, miscellaneous costs (i.e., tires, fuel and parts) can be difficult to investigate in detail. As such, equipment unit cost, namely hourly or daily equipment rate in dollars, becomes the major analytical criteria for most equipment. In some cases, such as transportation equipment, unit cost may refer to the rate by distance rather than by time. At the operational level, the equipment management team has to maintain a competitive equipment unit cost (rate) for two major reasons: 1) unit rates are applied to project estimating and bidding, which affects the bid success rate of the project management team; and 2) when the internal equipment rate is higher than that of a rental company, the project management team will seek the external resources (e.g. construction equipment), further lowering internal equipment use thus increasing unit cost.

The total cost and unit cost are convertible. To avoid confusion, this study focuses on the total (cumulative) cost of equipment, and the unit cost (rate) is determined based on the total cost and units worked (either in time or distance units), as illustrated in Eq. 1.1.

$$Unit\ Cost = \frac{Total\ Cost}{Units\ Worked} \tag{1.1}$$

Generally speaking, equipment total or unit costs are affected by ownership, maintenance, operating, and overhead costs; fleet size; equipment lifetime; equipment use; and other financial factors, such as interest rates. In current practice, it is widely accepted that two major analytical criteria need to be controlled by the equipment management team to maintain competitive equipment total or unit cost: equipment utilization and optimal fleet age ("sweet spot").

Equipment Utilization

Equipment utilization can represent one of several different concepts, depending on the management level at which it is applied. Utilization at a project level (El-Rayes and Moselhi 2001; Wang et al. 2004) is based on equipment downtime, which is affected by site and equipment conditions, operator skills, and force majeure (Prasad and Park 2004). Utilization at the corporate level assesses equipment use over its lifetime (Vorster 2009), and generally, high equipment utilization rates indicate efficient equipment use. At a corporate level, equipment utilization, or deployment (Vorster 2009), can be defined by Eq. 1.2.

$$Equipment\ Utilization = \frac{Used\ Time}{Total\ Ownership\ Time} \tag{1.2}$$

where *Total Ownership Time* is the time (in days or hours) that equipment has been owned by the corporation and *Used Time* is the time (in days or hours) that equipment is allocated to a project regardless of its operation on the jobsite.

Optimal Fleet Age ("Sweet Spot")

Optimal fleet age, also called the "sweet spot", is the analytical support for fleet replacement planning and serves as the basis for financial planning and capital budgeting decisions (Vorster 2009). Figure 1.2 shows the total cost per hour (orange), accumulated maintenance costs (black) and ownership cost (blue). The total cost per hour is the lowest at certain hours worked or service meter reading (SMR). The equipment hours worked to achieve the lowest cost per hour is the optimal fleet age, which is illustrated by the dashed line in Figure 1.2. Furthermore, the average age of the whole equipment fleet should be kept near the optimal fleet age. When the optimal fleet age is achieved, newer equipment and older equipment are well-balanced.

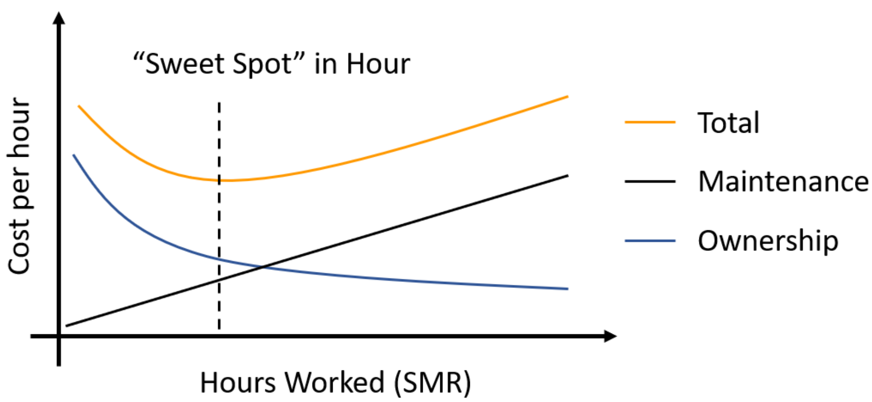


Figure 1.2 "Sweet Spot" Illustration

Optimal fleet age varies among equipment makes and models due to various factors such as the quality of the manufacturer. For a fleet of mixed equipment with assorted makes and models, the optimal fleet age can be difficult to calculate and achieve. Furthermore, optimal fleet age is a dynamic criterion due to nature of construction business. The supply and demand of equipment is changing every day; the balance can be achieved theoretically, but rarely practically.

Both "utilization" and "optimal fleet age" have been widely accepted and applied in construction equipment management and is easily determined from equipment management records. However, while large amounts of data are collected and updated dynamically for equipment management (e.g. new equipment transactions are made every day), current practice does not consider or use these data during the decision-making process. Simplified metrics are still applied, some of which may be calculated based on a small dataset. For example, to evaluate management performance, company decision-makers often rely on a single metric—equipment utilization. The use of simplified analytical tools may lead to missed opportunities for new management strategies based on already-collected data. Typically, construction companies with centralized construction equipment management strategies focus on routine construction operations and do not make good use of their existing data.

1.2 Problem Statement

The primary goal of this study was to introduce a new generation of data-driven analytics systems that will allow simulation-based analytics to be adjusted by dynamic or real-time data, thereby feeding dynamic equipment data into simulations to enhance equipment management processes. The proposed method or system could be applied to equipment management in equipment acquisition and disposal, shop operations, logistics, and fleet and asset management.

In the process of equipment acquisition and disposal, a large amount of data is collected. Trading construction equipment is a capital-intensive business process, and in the trading process, auctioneers and construction companies collect transaction records for many equipment makes and models. In the past, advanced predictive models (mathematical models, simulation models, and advanced models using artificial intelligence algorithms) have been proposed to assess the equipment's residual value, but two major challenges limit the application of these methods in practice: 1) the rigid nature of these approaches requires the mathematical or heuristic model to be redeveloped when the micro-economic conditions of a particular geographical market change—a difficult and time-consuming task for many practitioners, and 2) the mathematical complexity underlying many of the simulation models is difficult for practitioners to understand, much less apply. Because of the practical limitations, practitioners continue to rely on subjective experience to determine the optimal time, price, and location to purchase used equipment, despite the large amount of transactional data available.

A large amount of cost data is collected in the process of equipment maintenance, including preventative maintenance, repair and rebuilds. Maintenance and ownership costs have been studied and modelled numerically separately in previous research; however, investigations that combine both cost models to develop a mathematical model remain unsolved. The literature also indicates that the lack of maintenance data has prevented researchers from developing complicated models that incorporate both cost models. Furthermore, models generally represent occurred equipment costs by a single or a static set of values, interpreted directly from the cost model. In such cases, the risk and uncertainty in equipment management are hard to obtain and, consequently, affect the results of decision making.

Geo-location data for equipment is collected in the process of equipment tracking (equipment dispatch and logistics). Although these data are available to evaluate the performance of equipment logistics (i.e., dispatch), ensuring that projects are completed on time and on budget is still based on rates of equipment utilization. While utilization rate is an important metric for evaluating equipment-management performance, it does not take into consideration the logistical effort associated with equipment management at a corporate level. Inefficient deployment of equipment between worksites and equipment shops can increase logistics-associated effort and expenditures and reduce the amount of time equipment is available to work.

To address the limitations of past research and effectively use the collected equipment management data, simulation-based, data-driven analytics were proposed to enhance decision-making. Data for equipment transactions and tracking were investigated to provide advanced data-driven strategies. The following challenges were addressed:

- A quantification method was needed to evaluate the likelihood of achieving bargains in equipment trading.
- Equipment dispatch at the intra-organizational level was quantitatively studied to take advantage of equipment-tracking data.
- Novel performance metrics were proposed and validated to obtain a comprehensive evaluation of equipment-management performance.

- Bayesian inference was applied for dynamically updating analytical results to achieve an interpretable and user-friendly system.
- Simulation results were compared to analytical results through case studies to validate the proposed methods.

1.3 Research Objectives

The overall objective of this research was to develop data-driven approaches to enhance the equipment management practices of construction companies. This research created a dynamic, data-driven decision-support system to achieve the following objectives:

Objective 1 To provide analytical support to the decision-makers in equipment acquisition and disposal, especially for informing practitioners of ideal times, locations, and makes/models of equipment to purchase or sell. The following activities were undertaken to achieve this objective:

- Created advanced data-driven prediction models for bargain likelihood based on the equipment residual market value (RMV)
- Created dynamic (e.g. Bayesian inference-based) methods capable of integrating historical and dynamic data to more dependably predict the likelihood of acquiring equipment at bargain values (i.e., lower-than-market).

Objective 2 To quantify the life-cycle cost of heavy equipment, while incorporating both maintenance and ownership costs and to predict the cumulative total cost of equipment. The following activities were undertaken to achieve this objective:

- Applied K-means clustering and Expectation-Maximization algorithm to distinguish the maintenance stages of equipment, and to further generate distributions of these points.
- Applied simulation-based approached to quantify the uncertainties embedded in the equipment costs incorporating both ownership and maintenance cost.

Objective 3 To evaluate the logistical effort associated with equipment dispatch and planning, and to enhance equipment management through improved decision-making. The following activities were undertaken to achieve this objective:

- Proposed performance metrics by using the social network theory driven by the equipment logistics data.
- Demonstrated the feasibility of the proposed approach by using real equipment logistics data collected from the equipment and project management systems.

1.4 Research Methodology

The framework followed in this thesis used the concept of a dynamic data-driven application system (DDDAS) (Darema 2004). The DDDAS approach seeks to address the challenges of equipment management by incorporating real-time sensor data into running simulations. This framework can make use of various analytical methods, such as data mining, optimization algorithms, social network analysis, and simulations. The data-mining algorithms can be applied to identify missing data or outliers. Simulation methods can fill the gaps of multi-relational data and validate the proposed approach. Algorithms and models that facilitate analytics can use the transformed data with simulation models to generate selected metrics for a given decision-support application within the framework. This concept facilitates the dynamic addition of new data into simulation models to enhance the accuracy and predictability of the original models.

In this study, the concept of a DDDAS was implemented to develop a system using data from equipment trading, tracking and management systems, and project management systems. This specialized framework consisted of three major components, namely data inputs, data-driven analytics and data outputs to provide analytical decision support, as illustrated in Figure 1.3. The outputs can be categorized into three parts to support decision making and determine the strategies in equipment management: 1) equipment trading strategy, 2) equipment replacement strategy, and 3) equipment dispatch strategy. The output for the equipment trading strategy was the "likelihood of a bargain", which is the product of bargain percentage and transaction quantity, and provided analytical support to achieve bargains in equipment acquisition and disposal. For the equipment replace strategy, the analytical output was the "life-cycle cost of equipment", which assisted decision makers in understanding the uncertainties associated with the maintenance and ownership equipment costs. For the equipment dispatch strategy, the output was direct dispatch index (DDI), a novel performance metric that further evaluated the equipment logistics performance.

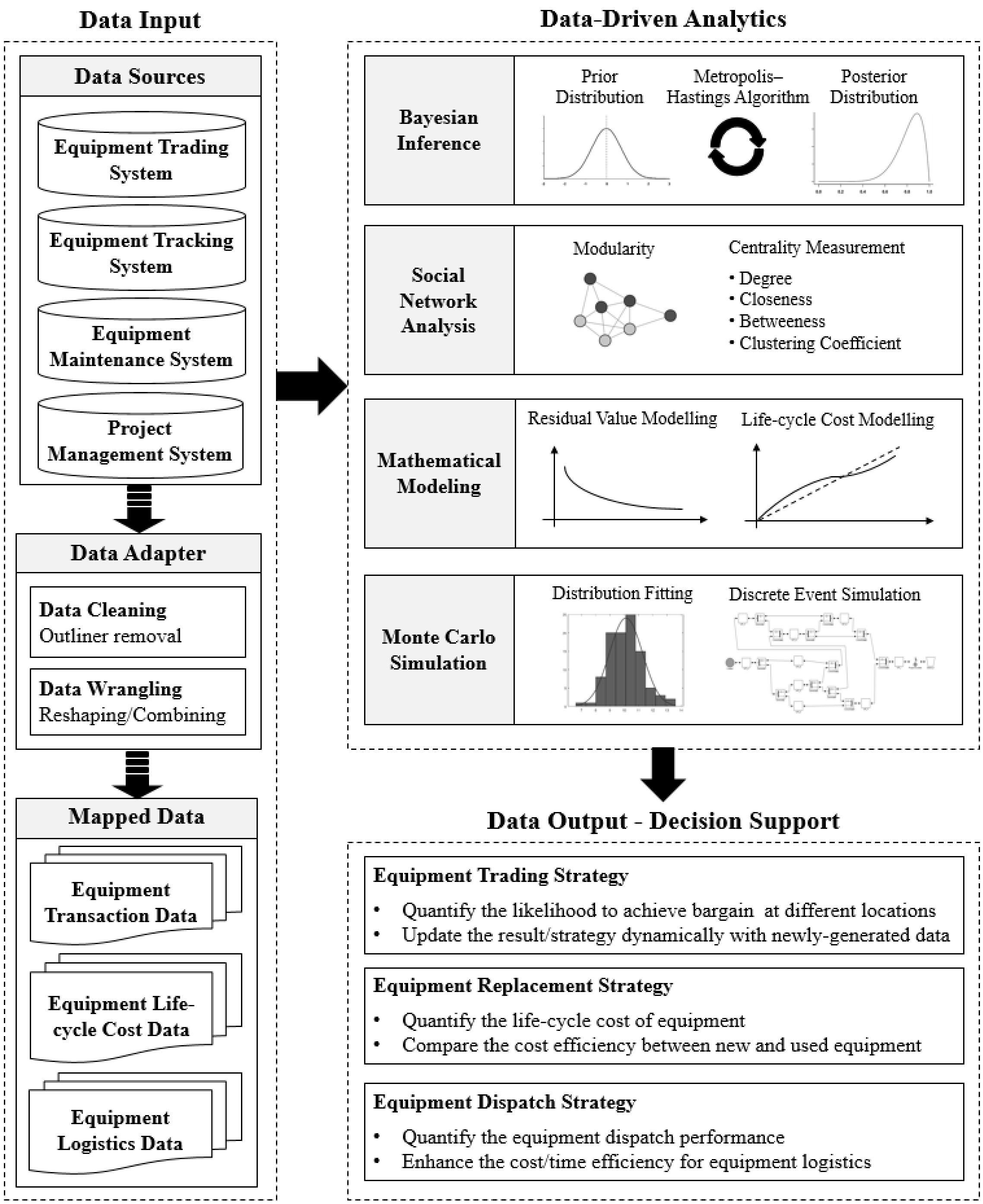


Figure 1.3 Flowchart of the Methodology to Obtain Data-Driven Strategies

1.5 Thesis Organization

This thesis is organized following a paper-based format that is consistent with the research framework shown in Figure 1.3, and the relationships between chapters are listed and explained in Table 1.1. Chapter 2 focuses on the methodology to achieve an equipment trading strategy, which could be used to obtain bargains in equipment acquisition and disposal. Chapter 3 discusses the development of equipment life-cycle cost that leads to the equipment replacement strategy based on both equipment ownership and maintenance cost. In Chapter 3, the ownership costs data studied in Chapter 2 was further investigated and new data, namely maintenance cost data, were added and studied with the ownership costs data. Chapters 4 focuses on the equipment logistics data and develops a novel analytical model to obtain decision-support metrics to evaluate the equipment logistics performance.

Table 1.1 Relationship between Chapters.

<table>
<tr><th>Data Input</th><th>Raw Data Sources</th><th colspan="2">Data-Driven Strategies</th></tr>
<tr><td>Ownership Costs</td><td>Equipment Management System
Equipment Auction System</td><td>Chapter 2: Equipment Trading Strategy</td><td rowspan="2">Chapter 3: Equipment Replacement Strategy</td></tr>
<tr><td>Maintenance Costs</td><td>Equipment Management System
Corporate Accounting System</td><td></td></tr>
<tr><td>Equipment Logistics</td><td>Project Management System
Equipment Management System
Equipment Tracking System</td><td colspan="2">Chapter 4: Equipment Dispatch Strategy</td></tr>
</table>

Detailed contents of each chapter are listed below.

Chapter 2: Dynamic, Data-Driven Decision-Support Approach for Construction Equipment Acquisition and Disposal

- Adapted a Bayesian inference-based approach to automatically update historical data with newly-generated market data.
- Conducted a case study to validate the feasibility and applicability of proposed solutions.
- Defined a novel "likelihood of bargain" concept for assisting practitioner's decision-making.
- Developed a user-friendly, easy-to-understand method for incorporating transaction information in real-time to sell or buy used construction equipment.

Chapter 3: Data-Driven Simulation-based Analytics to Assess the Life-cycle Cost of Equipment

- Compared the life-cycle cost of new equipment and used equipment
- Conducted a case study to validate the proposed model.
- Modeled the life-cycle cost of equipment mathematically.
- Quantified the uncertainties associated with the maintenance, ownership and total costs of the equipment over the equipment lifetime.

Chapter 4: Equipment Logistics Performance Measurement Using Data-Driven Social Network Analysis

- Conceptualized the equipment logistics problem into a social network model
- Conducted a case study to validate the feasibility and applicability of proposed solutions.
- Defined a novel performance metric for assessing equipment-logistics performance.
- Developed a data-driven approach to evaluate the logistical effort in equipment dispatch.
- Used social network analysis (SNA) to visualize equipment- and project-tracking data.

Chapter 5: A summary of the research contributions, limitations, and envisioned future work

2 CHAPTER 2: DYNAMIC, DATA-DRIVEN DECISION-SUPPORT APPROACH FOR CONSTRUCTION EQUIPMENT ACQUISITION AND DISPOSAL

2.1 Introduction

The successful completion of construction projects depends on reliable and functional equipment fleets (Vorster 2009). Equipment fleets are maintained and managed, in part, through the acquisition and disposal of used construction equipment. Construction companies aim to reduce costs by purchasing used equipment at rates that are lower than average market values.

These companies often assess equipment value using the residual market value (RMV), which is the estimated value of used construction equipment at a point in time (Lucko 2011). Notably, by dividing the RMV by a list price, the RMV can be further expressed as a residual value percentage (Vorster 2009; Zong 2017). Successful acquisition and disposal of construction equipment—a capital-intensive business process—requires both experience and expertise, particularly when assessing fair market values of equipment. While advanced predictive models capable of assessing equipment residual market values have been developed, these models cannot be automatically updated with new market data, rendering them less and less accurate over time.

Practitioners are increasingly looking to alternate geographical markets for equipment bargains, where, as a consequence of region-specific market factors, the price of equipment may be lower than average market rates. However, identification of geographical markets where bargains are likely is a difficult task that requires the consideration of various micro-economic factors together with historical transaction data. While various approaches designed to model the RMV of specific equipment types from recent transaction data and corresponding predictive models have been proposed, two major challenges limit the application of these methods in practice. Specifically, (1) the rigid nature of these approaches require the mathematical or heuristic model to be redeveloped when the micro-economic conditions of a particular geographical market change—a difficult and time-consuming task for many practitioners and (2) the mathematical complexity underlying many of the models makes it difficult for practitioners to understand, and in turn feel comfortable relying on the results of the model. Because of the practical limitations of these approaches, practitioners continue to rely on their subjective experience to determine the optimal time, price, and location to purchase used equipment despite the tremendous amount of quantitative transaction data available.

To address these limitations, this study has developed a user-friendly, easy-to-understand method for incorporating transaction information updated dynamically to more accurately predict the optimal time, location, and price to sell or buy used construction equipment. A Bayesian inference-based approach is used to automatically update historical data with newly-generated market data, alleviating the need for time-intensive redevelopments by end users. Furthermore, by applying the proposed method, practitioners are not required to understand complicated models. Additionally, a novel likelihood of bargain concept is introduced, allowing practitioners to make decisions based on the relative value of equipment rather than on absolute RMV. The functionality and validity of the proposed approach was demonstrated through the implementation in a practical case, where the method was found capable of accurately predicting which geographic locations were associated with the greatest likelihood of bargains.

2.2 Literature Review

Residual Market Value Prediction

Various multivariate mathematical models have been developed to predict RMV across a number of engineering disciplines. For example, the American Society of Agricultural Engineers (ASAE) has recommended a generalized regression formula to estimate the residual value percentage of agricultural and forestry equipment (ASAE 2003). Beginning with equipment age as the only factor in the formula (Bates et al. 1979), follow-up studies have since advocated for the inclusion of additional factors, such as horse power, manufacturer (Reid and Bradford 1983), and macro-economic indicators (Cross and Perry 1995).

In construction literature, a multi-linear regression RMV prediction model, based on historical auction sales data of various types and sizes of equipment, was first proposed (Lucko 2003). The RMV of construction equipment were later modeled as the second-order polynomial of equipment calendar age together with manufacturer, condition rating, auction region, and macro-economic indicators (Lucko and Mitchell 2010). Notably, Lucko and colleagues (2006) found that residual values often differed from transaction prices, which were also influenced by condition, market, and the macro-economy. Indeed, an empirical analysis of price relationships in a multi-item, multi-type auction by Ponnaluru (2009) determined that, while certain factors such as manufacturer, condition, region, and auctioneers had a significant impact on selling price, the influence of these factors was, at times, unpredictable.

With the development of data mining technology, automated modeling using an autoregressive tree algorithm was used to predict the RMV of wheel loaders (Fan et al. 2008), which was found to outperform artificial neural network (ANN) and multi-liner regression-based models in terms of accuracy. Recently, other data mining algorithms, such as K-nearest neighbor (KNN) and random forest (RF), outperformed the regression tree model in terms of computational accuracy and cycle time for predicting the RMV of articulated trucks (Zong 2017). Notably, the inputs were comprised of auction location, auction year, eleven factors describing the mechanical condition of the equipment, and two other factors regarding the macro-economy. Although determined to be more accurate, the application of data mining algorithms for RMV modeling and prediction has not become widespread in practice, as (1) practitioners find the algorithms and their outputs difficult to understand and interpret and (2) collecting and ensuring the accuracy of required model inputs is a time-consuming task.

Due to their simplicity, reliability, and efficiency, straight-line models and declining exponential models have emerged as two popular approaches used to estimate RMV in the construction industry (Vorster 2009). Notably, findings that RMV decline fastest in a machine’s early life indicate that RMV of construction equipment is best estimated using a declining exponential curve (Vorster 2009). Exponential regression models have been successfully used by the North Carolina Department of Transportation to represent the decline of RMV over machine age (Kauffmann et al. 2013), and the highest coefficient of determination has also been reached using an exponential regression function when assessing resale values of powertrains (Kleiner and Friedrich 2017). Considered a reliable and interpretable model that can be easily updated with new data over time, the exponential model with a single factor (i.e., age or engine hours) has become widely accepted and used in construction equipment management practice. Models and the number of input factors required to determine RMV are summarized in Table 2.1.

Table 2.1 Models and Number of Input Factors Required to Determine RMV

Year	Author	Model/Method	Number of Factors
2006	Lucko	Second-order polynomial multi-linear model	3
2008	Fan et al.	Autoregressive tree algorithm (ART)	9
2009	Vorster	Declining-exponential model	1
2017	Zong	K-nearest neighbors algorithm (KNN)	11

Transaction Quantity and Transportation Costs

Previous equipment management studies have focused on the development of methods capable of determining the optimal selling time of equipment based on residual prediction models (Bates et al. 1979; Mitchell 1998; Ponnaluru et al. 2012). As such, one major factor, transactions quantity at different ages, were missed in these studies. Residual value grid, as shown in Figure 2.1, was illustrated to show the age at which the equipment was sold and the actual transaction price and the number of transactions at a given age and price (Vorster 2009).

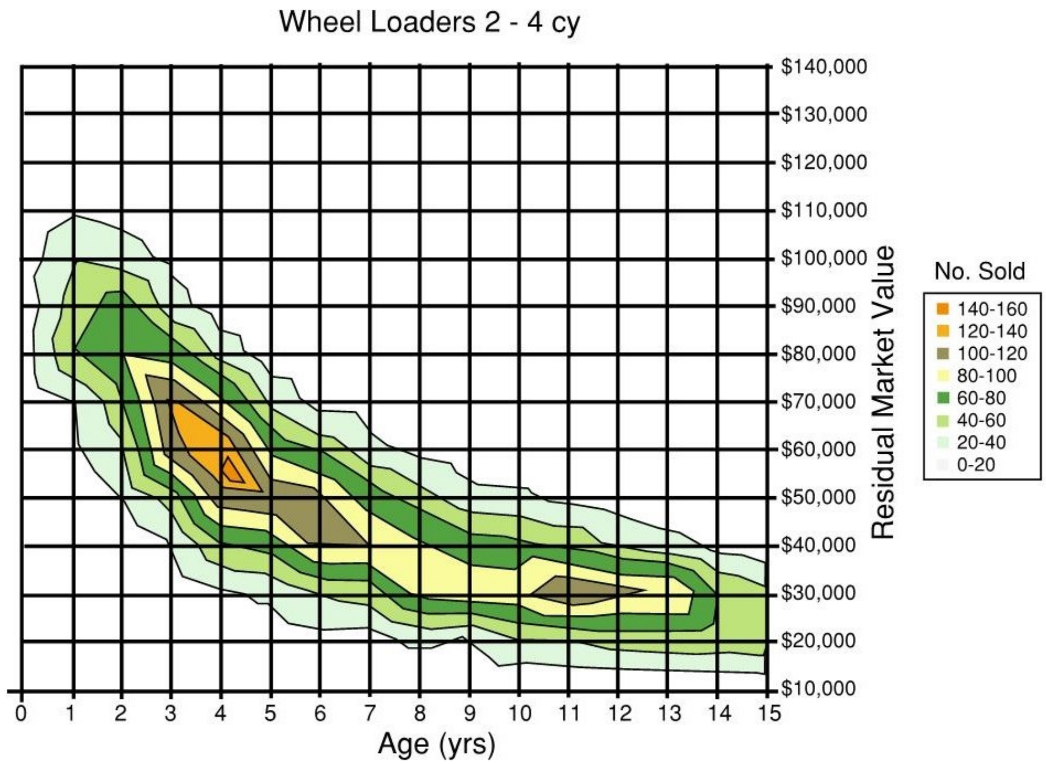


Figure 2.1 Residual Value Grid Example (Vorster 2009)

Since the range of transaction prices is large for the equipment sold at the same age, the buyers or sellers can benefit from purchase or sell at different locations. For example, to purchase a 5-year-old wheel loader in California, the average cost within the state is about $50,000 and the lowest cost from the other states could be $29,000 which is from Texas. Assuming a $6,000 transportation cost to transport equipment from Texas to California, the buyer can still save about $15,000 for purchasing similar used equipment. In short, the transportation costs are critical in finding the optimal time. In this study, the proposed approach will investigate the transaction quantities and transportation costs which may lead to the adjustments of existing equipment trading strategies.

Optimal selling time is dynamic, changing as the number of potential buyers and sellers fluctuate in the market. This is particularly important for construction equipment, where (1) a majority of transactions have been found to occur at the mid-point of an equipment's lifespan, indicating that there are fewer potential sellers and buyers for relatively new or old equipment (Vorster 2009), and (2) specialty types and models of equipment, for which there are few buyers/sellers, are prevalent in construction. The consideration of transaction quantity as a factor representing equipment supply and demand, however, was not factored in the aforementioned RMV prediction models. Another important factor, particularly when considering alternate geographical markets, is the transportation cost associated with transporting equipment from the purchase location to the jobsite. As with transaction quantity, transportation cost was also not included in the RMV prediction models.

Bayesian Inference Applications in Construction

Bayesian inference, also referred to as Bayesian updating or Bayes' theorem, is widely accepted and used in the field of construction engineering and management for statistically updating mathematical models as new information becomes available (Chung et al. 2006; Song and Eldin 2012; Zhang et al. 2015; Ji and AbouRizk 2017; Wang and Zhang 2018). In Bayesian inference, the posterior distribution (i.e., updated probability) is achieved by considering both a prior distribution (i.e., estimated or derived from historical data) and a likelihood function (i.e., derived from newly observed data) (Gelman et al. 2003).

Bayesian inference was first applied in the construction engineering management domain in the area of tunnel project simulation. Bayesian inference was applied to achieve the inputs of a simulation model (i.e., distributions), which integrated both subjective expert experience and objective sampling data (Chung et al. 2006). Later, Bayesian inference was successfully applied to more complex tunnel construction models and was found capable of handling other repetitive construction projects, such as earthmoving operations (Zhang et al. 2013). Models for predicting geological conditions, as a major risk factor in tunneling projects, have also been developed using Bayes' theorem, allowing continuous updates of the simulation model and, in turn, resulting in considerable improvements in project performance (Zhang et al. 2015; Werner et al. 2018).

Bayesian inference can be applied to the estimation or prediction of analytical criteria. With the integration of continuous input data, Bayesian inference can ensure that estimations or predictions remain up-to-date. Bayesian inference was used to incorporate new global positing system (GPS) data in a real-time simulation of earthmoving loading processes, resulting in improvements in look-ahead scheduling accuracy and reductions in average wait and delay times (Song and Eldin 2012). The method has also been applied to estimate the nonconforming rate of welding operators, enabling opportunities for quality and productivity improvements (Ji and AbouRizk 2017). A method for more accurately estimating risk probability for public-private-partnership (PPP) projects through a Bayesian inference-based integration of historical data and expert judgement has also been developed and validated (Zhao and Fu 2006; Wang and Zhang 2018).

2.3 Methodology

A dynamic, data-driven mathematical model that (1) adapts to continuously changing market conditions by incorporating data updated dynamically and (2) considers transaction quantities and transportation costs to more efficiently and accurately identify opportunities for equipment bargains is proposed. The workflow of the proposed methodology is illustrated Figure 2.2.

In the proposed methodology, historical and new data are first collected, cleaned, and consolidated, as required. Historical data are used to establish a baseline for RMV analysis (Figure 2.2; *solid arrow*), and new market data are used for updating purposes (Figure 2.2; *dashed arrow*). New and historical data are integrated using Bayesian inference, and updated bargain percentages and transaction quantities are determined. The likelihood of a bargain, from the perspective of a buyer, is then calculated as the product of the bargain percentage and transaction quantity.

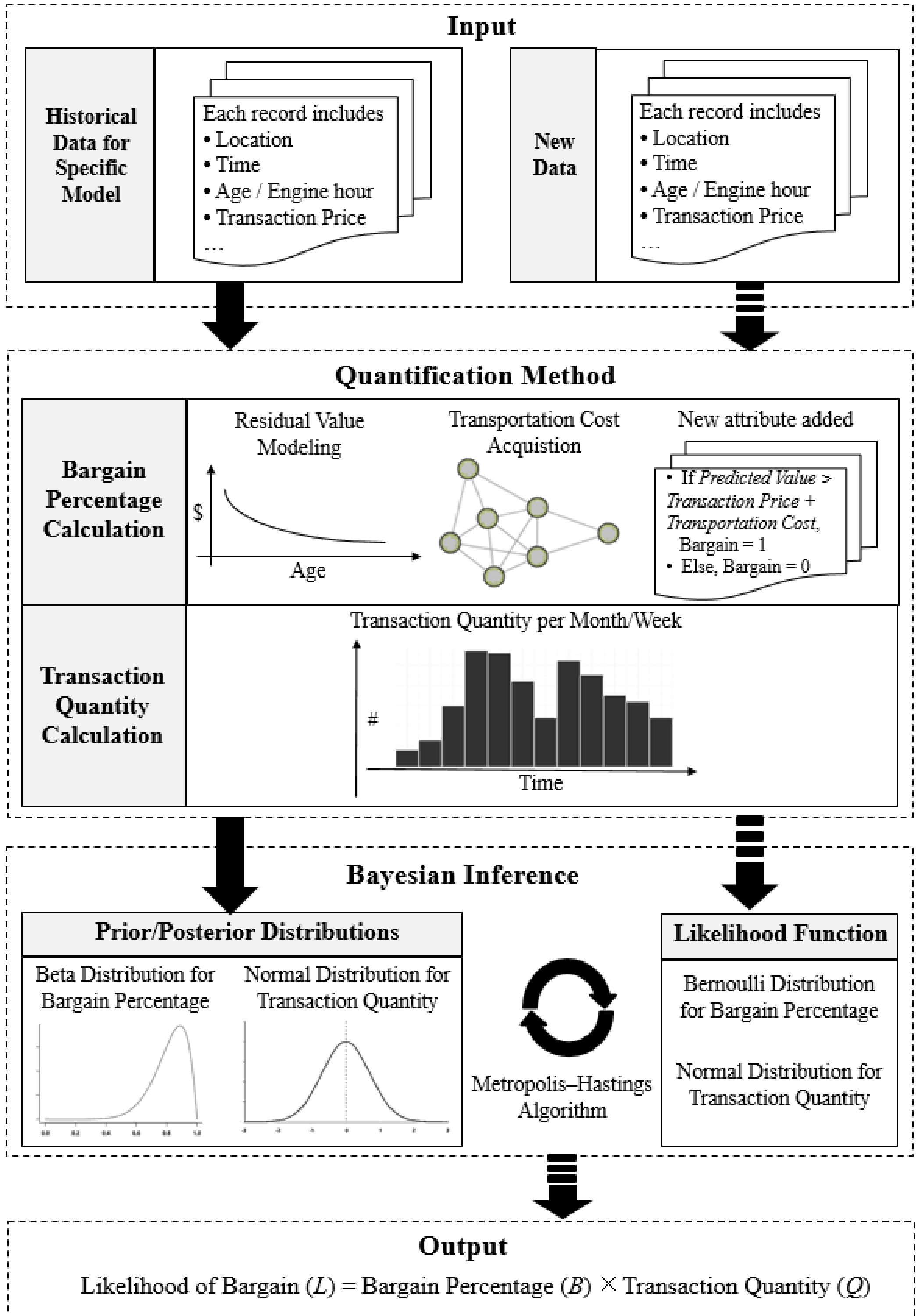


Figure 2.2 Proposed Flowchart of Dynamic Equipment Trading Assessment

Data Input

Input data can be collected from a variety of sources, such as from a company's internal accounting system, the database of an auctioneer, or a third-party data source, each with their own advantages and disadvantages. Although readily available, data stored in a company's accounting system are often pre-screened, processed, and edited by analysts, which may be inconsistent with the original data (Mitchell 1998). For example, a typographical error may alter the analytical results, and considerable data cleaning efforts are, therefore, required. International equipment auctioneers collect transaction data for corporate operations and fair-trading purposes, allowing potential equipment buyers to acquire these data free of charge. Third party, commercial software companies have capitalized on this, consolidating auctioneer data into large databases. Although subscription fees of these third-party data sources may be substantial, the data are comprehensive and consistent. Typically, these data cover the transaction records for various equipment makes and models, are collected from various sources, and include transaction information updated dynamically (i.e., within hours or days). It is important to note that, compared to data sources that are continuously updated, data from a company's internal systems are likely insufficient for the proposed method. However, company data may be used to enrich other data sources through data consolidation.

Prior to consolidation, data are cleaned as required; notably, the methods used and the extent to which the data are cleaned will depend on the characteristics of the individual dataset(s). In the proposed method, six factors are required as input, namely the date and location of the transaction as well as the manufacturer/model, service meter reading (SMR), and the selling price (in USD and CAD) of the equipment. Example data are summarized in Table 2.2.

Table 2.2 Sample Input Data

Date	Manufacturer/Model	SMR	Location	Price (USD)	Price (CAD)
09-Mar-18	Case 850L XLT	1755	Indiana	$82,000.00	$106,600.00
16-Oct-17	Case 850L XLT	2626	Alabama	$46,000.00	$59,800.00
06-May-15	Case 850L XLT	5168	Texas	$57,500.00	$75,700.00

Quantification Method

Residual Market Value

Although a number of quantification methods have been reported (Table 2.1), the proposed methodology has opted to use the exponential declining model that includes only the SMR as a factor (Lucko et al. 2007; Fan et al. 2008; Vorster 2009; Zong 2017). The rationale for selecting this model is based on (1) the predominant use of this model throughout literature and in practice, (2) the relative ease-of-use of this model compared to other methods, and (3) the focus of this research, which is on the dynamic updating of bargain likelihood as opposed to RMV determination. Nevertheless, it is important to note that the approach used to determine RMV in the proposed methodology is not limited to the exponential declining model and can be easily replaced by other mathematical models as desired.

The exponential model of RMV can be mathematically expressed as Eq. (2.1) and is calculated, based on available transaction data of a specific equipment model, using the least square method. The RMV is calculated, using Eq. (2.1), for each transaction record.

$$RMV = a \cdot e^{b \cdot SMR} + c \tag{2.1}$$

where RMV is the residual market value, SMR is the service engine reading in hours, and a, b, c are constants used to adjust the magnitude of RMV results.

Transportation Cost

Transportation costs are influenced by (1) the distance traveled between the location where the equipment is to be used and the location where it is purchased and (2) the transportation method, which encompasses transportation rate and time. In practice, transportation costs are most accurately determined from quotations obtained from transportation companies. However, a simplified approach for determining transportation cost may also be used, albeit with less accurate results. In this study, the transportation costs are calculated using Eq. (2.2), which is the product of hourly rate and travel time.

$$TC_i = r_i \times h_i \tag{2.2}$$

Where TC_i is the transportation cost, r_i is the hourly cost of transportation, and h_i is the total round-trip travel time for transaction record i. A conservative assumption that the time required for travel includes the return of the empty transport truck (i.e., round-trip) is applied in the present study; however, the assumed travel time can be modified as desired.

Bargain Factor

The calculated RMV is then compared to the sum of the transaction price and transportation cost, and a binary bargain factor is determined for each transaction record using Eq. (2.3).

$$B_i = \begin{cases} 0 & RMV_i < (TP_i + TC_i) \\ 1 & RMV_i \geq (TP_i + TC_i) \end{cases} \tag{2.3}$$

where B_i is the bargain factor, RMV_i is the residual market value calculated from Eq. (2.1), TP_i is the transaction price, and TC_i is the transportation cost of transaction record i.

Bargain Percentage

A bargain percentage, which represents the number of bargains per transactions for a defined period of time, is then calculated. Transaction quantity (i.e., the total number of transaction records) is calculated using Eq. (2.4) and, based on the transaction quantity and bargain factors, Eq. (2.5) is used to calculate the bargain percentage.

$$Q = n \tag{2.4}$$

$$B = \begin{cases} \frac{\sum B_i}{n} & n > 0 \\ 0 & n = 0 \end{cases} \tag{2.5}$$

where Q is the total number of transactions (equipment auctions), n is the total number of transaction records, B is the bargain percentage for all transaction records, and B_i is the bargain factor of transaction record i.

Bayesian Inference

In this study, Bayesian inference is used to update prior information (i.e., expert opinion, distribution derived from historical data, or a non-informative distribution), termed the prior distribution, with new beliefs or observations about the data (i.e., newly generated market data), termed the posterior distribution (Ji and AbouRizk 2017). In general, Bayesian inference derives the posterior distribution from two sources, the prior distribution and the likelihood function (Gelman et al. 2003), where the likelihood function is the function of the parameters in a statistical model that can be derived from new observations.

The posterior distributions for the two variables studied here, namely bargain percentage (B) and transaction quantity (Q), are calculated using Eq. (2.6) and Eq. (2.7).

$$P(B|A) = \frac{L(A|B) \cdot P(B)}{P(A)} \tag{2.6}$$

$$P(Q|P) = \frac{L(P|Q) \cdot P(Q)}{P(P)} \tag{2.7}$$

where $P(B|A)$ and $P(Q|P)$ are the posterior distributions of B and Q, respectively; $P(B)$ and $P(Q)$ are the prior distributions of B and Q, respectively; $L(A \mid B)$ is the likelihood function that provides the distribution of new data given the bargain percentage B; $L(P \mid Q)$ is the likelihood function that provides the distribution of new data given the transaction quantity Q; $P(A)$ is the marginal distribution of bargain percentage in the new data; and $P(P)$ is the marginal distribution of transaction quantity in the new data.

Specifically, taking bargain percentage as an example, $P(B)$ is the prior distribution which is based on the distribution derived from historical data for bargain percentage. $P(A)$ is the marginal distribution which is based on the distribution derived from new data. $P(B|A)$ is the posterior distribution integrating both the historical data and new data.

In this study, the prior distribution is achieved from the historical data where the distribution fitting methods are applied. To estimate the parameters of the selected probability distributions, the maximum likelihood estimation method is applied, which is a widely-used method (Law 2007). After obtaining the parameters, Anderson-Darling tests and visualization test are conducted to ensure the goodness of fit (Biller and Gunes 2010).

That the product of a valid prior distribution and likelihood function are integratable is not certain (Fink 1997), as it is not always possible to find an analytical solution to a well-defined integration problem in Bayesian inference. Adding further complexity, the prior distributions and the poster distributions are often not in the same probability distribution family. For example, a prior distribution may be uniform (as may be case when no prior information is available), while the posterior distribution after Bayesian inference may be beta or normal. In this case, it is difficult to control generated posterior distribution(s), resulting in unstable model performance.

In contrast, a conjugate distribution refers to a special case in Bayesian inference when the posterior and prior distributions belong to the same probability distribution family (Fink 1997). Accordingly, conjugate distributions are a means of solving high-dimensional problems that are computationally infeasible, as updates are made to the distribution parameters instead of the complex integrals.

Transformation of conjugate distributions (from prior to posterior distributions) have been formalized and well-studied (Raiffa and Schlaifer 2000; Fink 1997). For the bargain percentage *(B)*, the inferring variable is either 0 or 1 obtained from the Eq. (2.3) which forms a Bernoulli distribution, namely binomial distribution. It is common to use beta distributions as the standard conjugate priors for inferring variable in a binomial distribution (Berger 1985). In addition, the primary reasons for choosing beta distribution for bargain percentage are (1) bargain percentage is bounded with the range of 0 to 1 (i.e. the hypothesis in using beta distribution as conjugate distributions can be accepted); (2) the flexibility of beta distribution can provide accurate outputs. For the transaction quantity (*Q*), the reason for choosing normal distribution is the parameters of normal distribution are intuitively meaningful and easy to estimate from the data. In addition, historical data of transaction quantity can fit normal distribution well based on Anderson-Darling tests and visualization test.

Two commonly-used conjugate distributions, the beta and normal distribution, can be applied to the bargain percentage and transaction quantity as summarized in Table 2.3. As such, bargain percentage and transaction quantity can be updated dynamically using Bayesian inference.

Table 2.3 Conjugate Distributions Used for Bayesian Inference

Factor Description	Factor Range	Likelihood Model	Conjugate Distributions	Distribution Range
Bargain Percentage (B)	[0,1]	Bernoulli	Beta (α, β)	[0,1]
Transaction Quantity (Q)	[0,+∞]	Normal with known variance σ^2	Normal (μ_0, σ_0^2)	[-∞,+∞]

Since direct sampling may not be achieved, and calculation of a normalization factor is computationally difficult, the Metropolis-Hasting algorithm—a Markov chain Monte Carlo (MCMC) method to approximate the distribution—is then applied. The Metropolis-Hasting algorithm was initially developed in 1953 (Metropolis et al. 1953), was extended to more generic cases in 1970 (Hasting 1970), and has since been used to solve similar, conjugate beta distribution problems in construction engineering and management (Ji and AbouRizk 2017). Briefly, the Metropolis-Hasting algorithm, associated with a target density f, requires the choice of a conditional density q, which produces a Markov chain $X^{(t)}$ through the following transition step (Robert and Casella 1999):

Given $X^{(t)} = x^{(t)}$,

1. Generate $Y_t \sim q\left(y|x^{(t)}\right)$.
2. Take

$$X^{(t+1)} = \begin{cases} Y_t & \text{with probability } \rho\left(x^{(t)}, Y_t\right), \\ x^{(t)} & \text{with probability } 1 - \rho(x^{(t)}, Y_t), \end{cases}$$

where

$$\rho(x,y) = \min\left\{\frac{f(y)}{f(x)}\frac{q(x|y)}{q(y|x)},1\right\}$$

In general, y refers to the new observation while $x^{(t)}$ refers to the prior knowledge. In this study, Metropolis-Hastings algorithm constructs two Markov chains of fraction nonconforming values of bargain percentage B and bargain likelihood L, namely $B^{(t)}\sim\{b^{(1)}, b^{(2)}, b^{(3)},\ldots\}$ and $L^{(t)}\sim\{l^{(1)}, l^{(2)}, l^{(3)},\ldots\}$. For bargain percentage B, given new observation y_B and the prior knowledge $b^{(t)}$, the Bayesian posterior $b^{(t+1)}$ will be accepted at a ratio $\rho\left(b^{(t)}, y_B\right)$. Similarly, for bargain likelihood L, given new observation y_L and the prior knowledge $l^{(t)}$, the Bayesian posterior $l^{(t+1)}$ will be accepted at a ratio $\rho\left(l^{(t)}, y_l\right)$.

Due to its built-in library and functions, the implementation of the Metropolis-Hasting algorithm is programmed in *R,* which is a popular, open-source software for statistical computing and graphics (R Core Team 2019).

Data Output

Outputs of the Bayesian inference are posterior distributions for the bargain percentage B (i.e., beta distribution, Table 2.3) and transaction quantity Q (i.e., normal distribution, Table 2.3). The bargain likelihood L is the product of the means for both posterior distributions. The decision to decompose bargain likelihood into two portions was based on mirroring the decomposition of risk into probability and severity in risk management (Aven 2016) to avoid the bias that occurs when using a fixed number to represent the magnitude of such situations. For example, in risk management, an incident with higher severity does not imply greater risk; similarly, in equipment management, a greater transaction quantity does not imply the market is in a “bargain” state. Bargain likelihood is calculated using Eq. (2.8).

$$L = \bar{B} \times \bar{Q} = \frac{\alpha_B}{\alpha_B + \beta_B} \times \mu_Q \tag{2.8}$$

where $\bar{B}$ is the mean value of the beta distribution for bargain percentage B, $\bar{Q}$ is the mean value of the beta distribution for the transaction quantity Q, α_B and β_B are two parameters modeling the beta distribution of bargain percentage B, μ_Q is the mean value of the normal distribution of transaction quantity Q.

In this study, the bargain likelihood is represented by a mean value but can also be defined as a distribution, which is still the product of both posterior distributions. As such, the distribution of bargain likelihood cannot be directly achieved through the proposed Eq. (2.8). Instead, the Monte Carlo simulation (explained in details through Appendix A) needs be applied to obtain the distribution of bargain likelihood through iterations. In each iteration, two random numbers are generated from both posterior distributions (i.e., beta distribution for bargain percentage, and normal distribution for transaction quantity, Table 2.3) respectively and, one randomized value of bargain likelihood can be achieved as the product of two random numbers. Upon iterations, given the posterior distributions of bargain percentage and transaction quantity, randomized values of bargain likelihood can be collected to form the distribution of bargain likelihood. According to this distribution, the uncertainty of bargain likelihood can be quantified along with its mean value.

Notably, bargain likelihood can be calculated for different locations or equipment models, which can then be compared with one another. The functionality of the proposed method and the potential decision-support insights that can be gleaned from its application are detailed in the implementation of proposed method that follows.

2.4 Implementation

Practical Case

To demonstrate the proposed dynamic, data-driven approach in detail, the method is applied to examine the bargain likelihood of the Caterpillar (CAT) 320 Hydraulic Excavator, which is widely used in the North American heavy civil construction industry because of their efficiency and fuel economy. Data used in the case were retrieved from a third-party, commercial system subscribed to by our research partner, a major construction contractor in Alberta, Canada. Historical transaction data for the period between 2012 and 2018 were extracted; 2,566 records from 52 states and provinces in both Canada and the United States were collected. Transportation cost of CAT 320 between states and provinces is achieved based on round-trip travel time (hr) and assumed hourly cost for equipment transportation ($120/hr).

To demonstrate the proposed method, historical data and new market data are required. Typically, during practical implementation, all available data will be categorized as historical data with new data, acquired by week, month, or year, as desired, used for updating purposes. However, the quantity of new market data available post-data extraction was insufficient to support the case and provide a solid demonstration. Therefore, for purposes of the practical case demonstration, transaction records were divided into two datasets representing historical and new market data. Transaction records between 2012 and 2014 were categorized as historical data, and the remaining data were considered as new data and used for updating purposes.

A declining exponential model, based on the historical data (Figure 2.3) and the least squares method, was developed to determine RMV. The declining exponential model, which describes the relationship between RMV and SMR is calculated as Eq. (2.9) and illustrated as the curve in Figure 2.3.

$$RMV = 106138 \cdot e^{-0.000128 \cdot SMR} + 26097 \tag{2.9}$$

where RMV is residual value of CAT 320, and SMR is the service engine reading in hours.

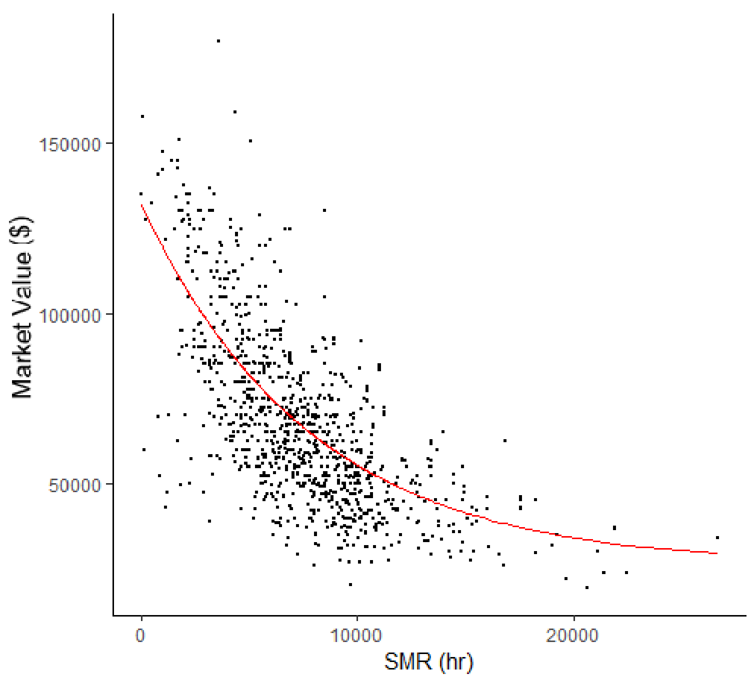


Figure 2.3 Scatterplot of Historical Data for CAT 320

The transportation is achieved based on the Eq. (2.10) and the round-trip travel time

$$TC_{ij} = 120 \cdot h_{ij} \tag{2.10}$$

where TC_{ij} is the transportation cost to move one machine (CAT 320) between location i and location j, and h_i is the total round-trip travel time between location i and location j.

Transaction records from Florida, USA, are used to illustrate the application of the proposed methodology. A mean bargain percentage $(\bar{B})$ and mean transaction quantity $(\bar{Q})$ were first determined. New market data were divided into three-month sections (2015Q1 to 2017Q4), and Bayesian inference was applied to update the means a total of 12 times, as listed in Table 2.4. Bargain likelihood (L) was calculated as the product of the two means using Eq. (2.8).

Table 2.4 Outputs of the Proposed Approach for Florida, USA

Time	Mean Bargain Percentage ($\bar{B}$)	Mean Transaction Quantity ($\bar{Q}$)	Bargain likelihood (L)
Historical	0.157	2.013	0.315
2015Q1	0.167	2.257	0.378
2015Q2	0.171	2.246	0.383
2015Q3	0.175	2.224	0.390
2015Q4	0.178	2.223	0.395
2016Q1	0.181	2.496	0.452
2016Q2	0.184	2.466	0.455
2016Q3	0.184	2.505	0.461
2016Q4	0.184	2.528	0.464
2017Q1	0.186	2.713	0.505
2017Q2	0.190	2.681	0.509
2017Q3	0.193	2.607	0.503
2017Q4	0.196	2.594	0.509

For the historical data, the bargain percentage and transaction quantity distributions are calculated using the method described in the previous "Quantification Method" section and are then updated as "new" data become available. As this practical case is focusing on comparing the bargain likelihood of various geographical markets, the transaction records were further divided into 52 geographical areas by state or province.

It is important to note, however, that the transaction records can be divided as desired. The proposed methodology was then applied to the transaction records of the 51 other states and provinces. The bargain likelihoods of the 10 states and provinces where equipment transactions were most frequent are listed in Table 2.5. Bargain likelihood trends are illustrated in Figure 2.4. Notably, bargain likelihoods were found to change over time. The 2nd quarter of 2017, indicated as dashed line in Figure 2.4 (*T1*) is illustrated as a map in Figure 2.5.

Table 2.5 Bargain Likelihood for 10 States and Provinces over Time

Time	TX	FL	ON	QC	CA	TN	AB	NC	SC	NM
Historical	0.406	0.315	0.196	0.180	0.174	0.099	0.078	0.075	0.074	0.074
2015Q1	0.478	0.378	0.207	0.171	0.184	0.116	0.096	0.093	0.088	0.077
2015Q2	0.494	0.383	0.195	0.163	0.19	0.128	0.117	0.106	0.089	0.07
2015Q3	0.506	0.390	0.194	0.176	0.193	0.137	0.130	0.118	0.089	0.071
2015Q4	0.505	0.395	0.185	0.186	0.204	0.164	0.140	0.112	0.089	0.077
2016Q1	0.516	0.452	0.196	0.196	0.181	0.158	0.133	0.125	0.088	0.077
2016Q2	0.539	0.455	0.202	0.204	0.186	0.166	0.145	0.130	0.088	0.071
2016Q3	0.547	0.461	0.197	0.212	0.205	0.162	0.142	0.141	0.089	0.071
2016Q4	0.563	0.464	0.190	0.199	0.195	0.168	0.137	0.149	0.087	0.075
2017Q1	0.562	0.505	0.192	0.205	0.202	0.172	0.134	0.156	0.087	0.076
2017Q2	0.563	0.509	0.194	0.214	0.198	0.176	0.130	0.162	0.088	0.072
2017Q3	0.571	0.503	0.189	0.220	0.200	0.186	0.139	0.171	0.088	0.073
2017Q4	0.560	0.509	0.177	0.214	0.207	0.175	0.137	0.174	0.088	0.073

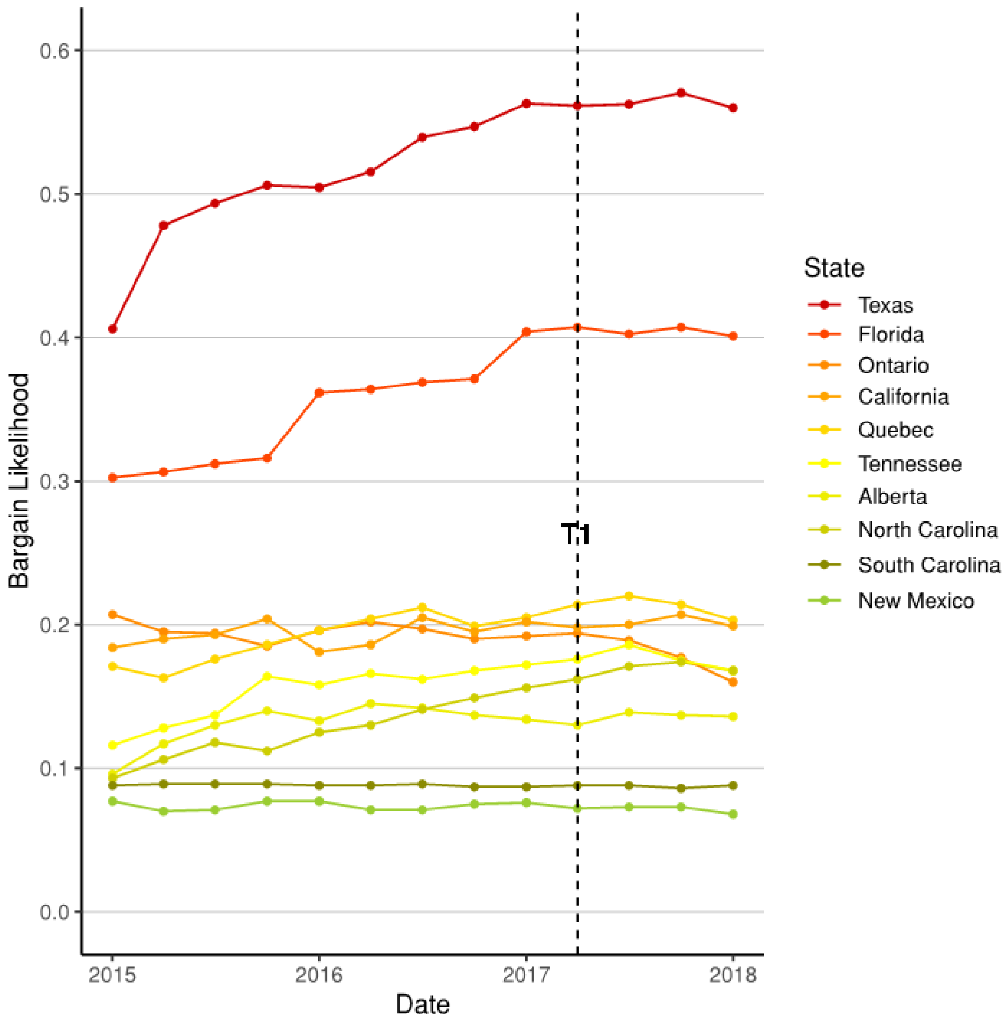


Figure 2.4 Bargain Likelihood Trends for 10 States and Provinces

By visualizing the bargain likelihoods of various geographical markets, as in Figure 2.5, allows decision makers to focus their time and resources exploring equipment purchasing opportunities in states or provinces where bargains are more likely to occur. In this particular case, for the 2nd quarter of 2017, bargains were predicted to have the greatest likelihood of occurring in Texas, California, Florida, Ontario, and Quebec.

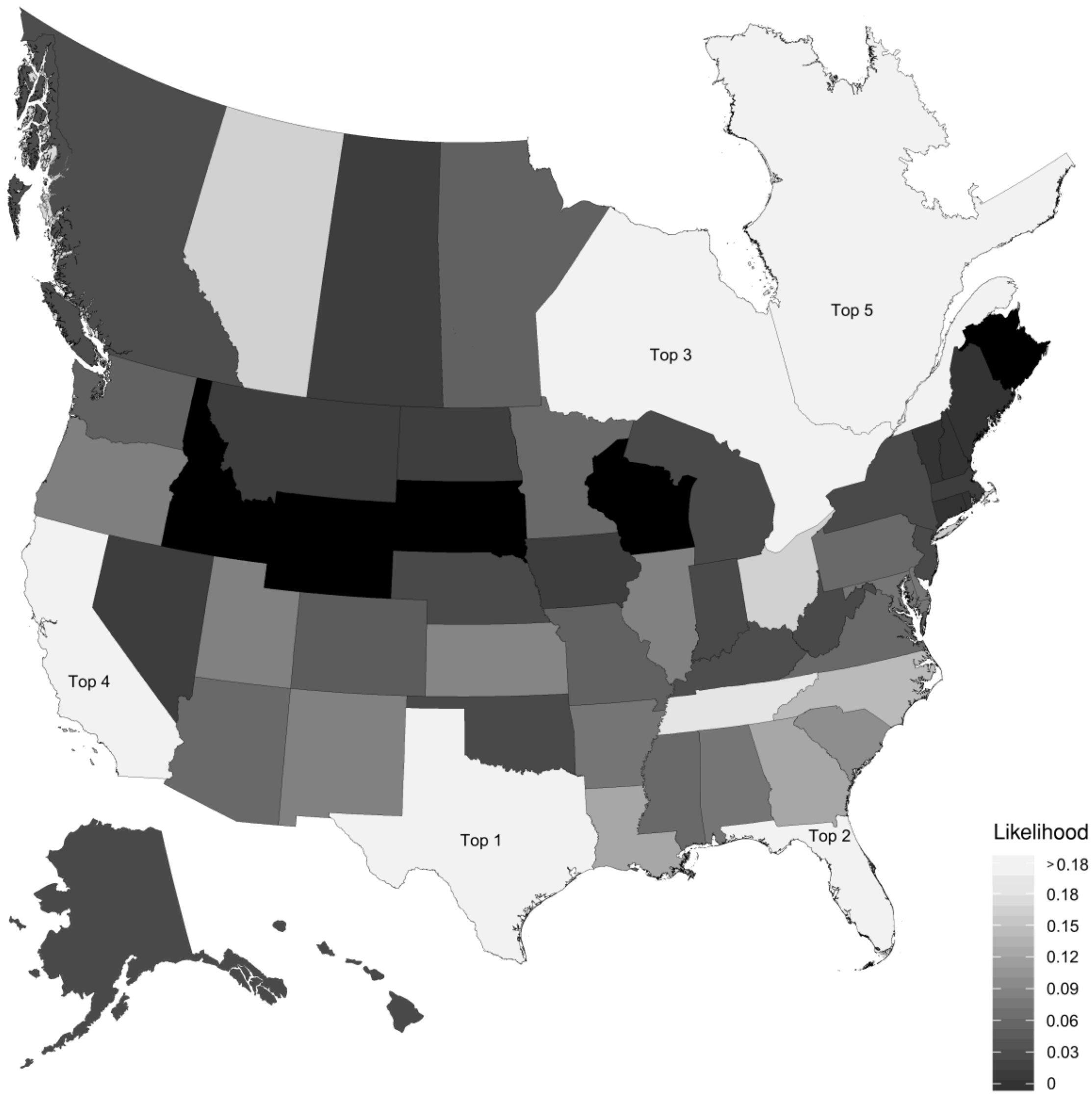


Figure 2.5 Bargain Likelihoods for 2nd Quarter of 2017 by Geographic Region

Discussion

The final transaction prices versus the SMR of all 2,566 transaction records for the CAT 320 were plotted in Figure 2.6. High and low point densities were represented as green and red, respectively. Similar to the work of Vorster (2009), this study found that (1) the range in transaction prices for equipment with similar SMR was large, indicating that RMV should only be used as a benchmark for equipment purchase or disposal and (2) most trading occurred at the mid-point of the equipment's lifespan, indicating that there are fewer sellers and buyers for relatively newer or older equipment.

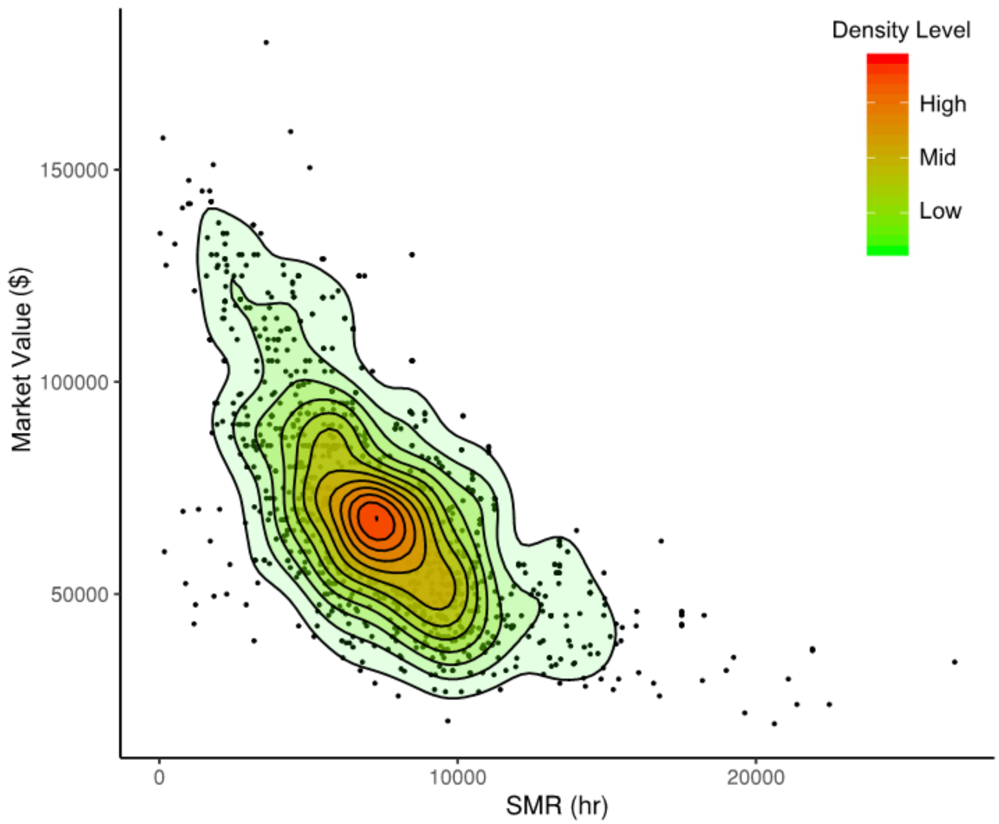


Figure 2.6 Density Plot of Transactions

The relationships between historical and future bargain numbers were also examined. Correlations between the number of bargains for the previous month (r^2=0.19) and the previous year (r^2=0.0076) versus the number of bargains in the next month are illustrated in Figure 2.7. A poor association between previous and future bargains was observed, indicating that practitioners cannot predict future bargains simply based on historical records and should, instead, rely on dynamically newly updated results, as introduced in this paper.

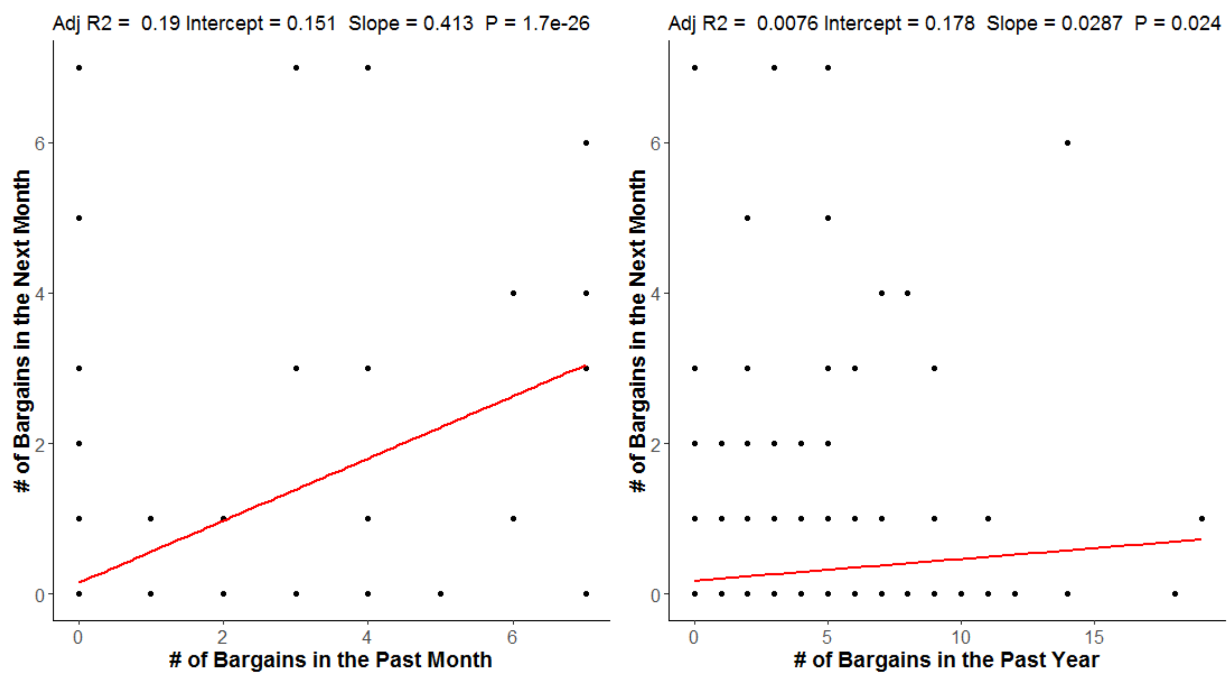


Figure 2.7 Correlation between the Past and Future Numbers of Bargains

The actual number of bargains after the prediction of likelihood is the primary criterion for method validation. Correlations between the bargain likelihood predicted by the proposed approach and the actual number of bargains in the future was used as a means of validating the proposed methodology. Correlations between bargain likelihood and the number of actual bargains for the next month (r^2=0.56) and the next quarter (r^2=0.70) are illustrated in Figure 2.8. A high correlation between bargain likelihood and the number of bargains, either monthly or quarterly, was observed, indicating that the proposed method was capable of dependably predicting areas where the number of bargains was high.

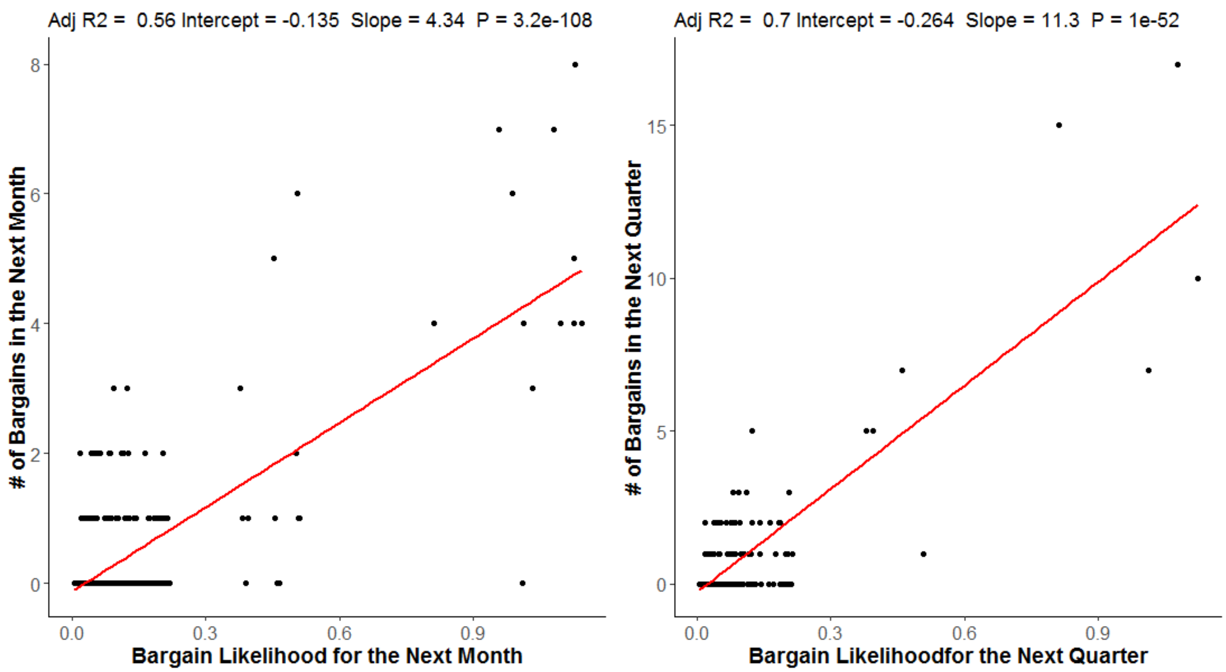


Figure 2.8 Correlation between the Number of Bargains and Bargain Likelihoods

Potential Adjustments for Implementations

For a generic implementation of proposed method, the adjustments can be made to the individual quantification module within the framework of methodology as illustrated in Figure 2.2. In other words, quantification modules may be replaceable with advanced methods, especially for the residual value modeling and transportation cost acquisition.

In the practical case, an exponential declining modelling is selected for RMV as demonstrated in Eq. (2.9). The comparison is made among the exponential model and aforementioned heuristic algorithms (Fan et al. 2008; Zong 2017). In these two research, heuristic models proposed for wheel loader include 9 factors (make, model, horsepower and etc.) while the models proposed for articulated trucks include 11 factors (machine age, make, model and etc.). RSE evaluates the percentage of the total squared error between the predicted value and actual value out of the total squared error if using the average of actual values as a prediction, which means low value of RSE means is ideal for modeling (Fan et al. 2008). As such, relative square error (RSE) and root relative squared error (RRSE) for 7 models are calculated to examine goodness of fit as summarized in Table 2.6.

Table 2.6 RSE and RRSE of the Models to Determine the RMV

Residual Value Modeling	Type/Make/Model	Data Size	RSE (%)	RRSE (%)
M5P (Zong 2017)	Articulated Trucks-Mixed	3,044	**12.9**	35.9
KNN (Zong 2017)	Articulated Trucks-Mixed	3,044	**11.6**	34.1
RF (Zong 2017)	Articulated Trucks-Mixed	3,044	**8.1**	28.5
Exponential Model (Eq.2.9)	Excavator CAT 320	2,566	1.4	11.8
ART (Fan et al. 2008)	Wheel Loaders-Mixed	8,589	5.5	23.3
ANN (Fan et al. 2008)	Wheel Loaders-Mixed	8,589	22.3	47.2
MLE (Fan et al. 2008)	Wheel Loaders-Mixed	8,589	40.4	63.6

Exponential declining model proposed in this study outperforms the models proposed for wheel loaders but perform worse than models for trucks, while all the models are good in fitness. In short, the exponential model can achieve accurate estimate of the residual value for the quantification module of proposed method. Furthermore, in the implementation of proposed method for other type, make or model of equipment, residual value modeling can be replaced with other mathematical model (liner or non-liner) or heuristic model. To generalize the proposed methodology framework, the selection of residual value model can made by the practitioners. It is highly suggested to compare the selected model with the previously proposed models based on RSE and RRSE to ensure the accuracy in RMV estimate.

Transportation cost to move construction equipment site by site is downplayed in this study. In the real world, transportation cost is determined by many factors including the remoteness of destinations, the weight of equipment and the availability of transportation companies. Two potential methods to replace the current method for transportation cost acquisition are summarized as follows: (1) transportation cost can be achieved based on quotations from transportation companies which are collected internally by other departments in the routine work such as project estimating and equipment dispatch; or (2) commercial systems/platforms (e.g. *uShip*) are specialized in facilitating equipment transportations and collecting data through users which brings the chance to construction companies in extracting valuable info. Similar to the RMV modeling, the acquisition of transportation costs can be selected by the practitioners when applying the proposed method into the real practice.

2.5 Conclusions

While purchasing equipment from remote markets involves transportation costs that impact profitability, local markets may be characterized by economic conditions resulting in equipment shortage and inflated prices. To capitalize on bargain opportunities, companies must not only have an understanding how much equipment is worth, but also how much it is likely to be sold for in other markets, particularly in auction-based settings where the final transaction price of equipment is unknown. Previous research has focused on developing a variety of methods capable of assessing RMV of equipment using historical equipment and transaction data. However, past efforts have not focused on developing methods that are able to dynamically and continuously adjust to new market data, rendering such models less and less accurate over time.

This research has proposed a dynamic, mathematical approach capable of addressing this academic gap. Two major parameters used in the proposed approach, namely bargain percentage and transaction quantity, are first defined. A detailed quantification method used to calculate RMV, transportation cost, and, ultimately, bargain percentage is established. Then, Bayesian inference is used to dynamically update models and parameters. In this study, a numerical solution is achieved through the implementation of the Metropolis-Hasting algorithm, which is applied to the two pairs of conjugate distributions defined for the bargain percentage and transaction quantity. Finally, the bargain likelihood, defined here as the product of bargain percentage and transaction quantity, is used evaluate the likelihood of observing bargains for a defined time period.

The functionality and validity of the proposed method was demonstrated following its implementation to a practical case, which examined transaction data for a popular hydraulic excavator, the CAT 320. Bargain percentage and transaction quantities were calculated from a "historical" portion of the data and were updated using "new" market data. Results (i.e., bargain likelihoods) obtained using the proposed method were found to strongly correlate with the actual number of bargains observed, indicating that the method was capable of generating valid, representative results.

Potential Applications for Decision-Support

The proposed method provides a practical advantage for decision makers, who can now adjust equipment acquisition or disposal strategies based on dynamic bargain likelihoods. By determining which geographic markets are most likely to have bargains available, practitioners can allocate resources and time to exploring potential bargain opportunities (i.e., auctions, used equipment online sales) in these specific locations. Similar to the comparison of geographic locations, the method can be used to compare various equipment makes/models to determine which makes/models are associated with greatest bargain likelihood. The method can also be used to examine bargain likelihood trends over time, allowing practitioners to plan to acquire or disposed of used equipment at times historically associated with bargain likelihoods. It is also important to note that the method, although discussed here from a purchasing perspective, can also be used to inform practitioners of bargain opportunities when selling used equipment.

Limitations and Future Work

The results of this study must be evaluated in consideration of certain limitations and challenges, which may be addressed in future work.

Acquisition of dynamic data requires access to multiple data sources. Currently, practitioners may access auctioneer data free of charge through individual auctioneers' websites. However, manually collecting and merging these data is tedious and inefficient. Alternatively, a contractor may opt to purchase a subscription to a commercial system, which compiles the data in a manner that is ready for use. The trade-off between the time required for and the cost associated with data collection should be evaluated by the practitioners.

RMV is fundamental to this study. As mentioned previously, other algorithms or models for determining the RMV can be embedded into the approach and tested in the future. Of particular interest would be the comparison of results obtained using heuristic algorithms, such as the artificial neural network, with those obtained using the declining exponential model.

The accurate determination of transportation costs is critical for obtaining dependable results. Many construction organizations rely on equipment transportation companies for equipment dispatch and logistics. As such, transportation costs can, and should, be collected and updated as part of the proposed process.

2.6 Acknowledgements

This work was generously supported by Graham Industrial Services LP and was funded by a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Science and Engineering Research Council of Canada (NSERC). The authors also would like to acknowledge Dr. Catherine Pretzlaw for her assistance with manuscript editing and composition.

2.7 References

American Society of Agricultural Engineers (ASAE). (2003). “Agricultural machinery management data.” *ASAE D497.4*, American Society of Agriculture and Biological Engineers, St. Joseph, MI.

Aven, T. (2016). “Risk assessment and risk management: Review of recent advances on their foundation.” *European Journal of Operational Research*, 253(1), 1-13.

Bates, J., Rayner, A., and Custance, P. (1979). “Inflation and farm tractor replacement in the US: a simulation model.” *American Journal of Agricultural Economics*, 61(2), 331.

Biller, B. and Gunes, C. (2010). "Introduction to simulation input modeling." *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, Maryland, USA

Chung, T., Mohamed, Y., and AbouRizk, S. (2006). “Bayesian updating application into simulation in the NEST tunnel project.” *Journal of Construction Engineering and Management*, 132(8), 882–894.

Cross, T. L., and Perry, G. M. (1995). “Depreciation patterns for agricultural machinery.” *American Journal of Agricultural Economics*, 77(1), 194-204.

Fan, H., AbouRizk, S., Kim, H., and Zaïane, O. (2008). “Assessing residual value of heavy construction equipment using predictive data mining model.” *Journal of Computing in Civil Engineering,* 22(3), 181-191.

Fink, D. (1997). “A compendium of conjugate priors.” <https://arxiv.org/pdf/1703.01444.pdf> (February 11, 2019).

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2003). *Bayesian data analysis*, Chapman and Hall, London, UK.

Hastings, W.K. (1970). “Monte Carlo sampling methods using Markov chains and their applications.” *Biometrika*, 57(1), 97-109.

Ji, W., and AbouRizk, S. M. (2017). “Credible interval estimation for fraction nonconforming: Analytical and numerical solutions.” *Automation in Construction*, 83(2017), 56-67.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*, Springer, New York, NY.

Kauffmann, P., Williams, R., and Hildreth, J. (2013). “Fleet Management Performance Monitoring.” <https://connect.ncdot.gov/projects/research/RNAProjDocs/2012-07FinalReport.pdf> (February 11, 2019).

Kleiner, F., and Friedrich, H.E. (2017). “Maintenance and repair cost calculation and assessment of resale value for different alternative commercial vehicle powertrain technologies.” *Proceedings of the EVS30 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium*, European Association for Electromobility, Brussels, Belgium, 1-16.

Law, A. M. (2007). *Simulation Modeling and Analysis*. McGraw-Hill Education. New York City, NY.

Lucko, G. (2003). *A statistical analysis and model of the residual value of different types of heavy construction equipment*, Doctoral Dissertation. Virginia Institute of Technology, Blacksburg, VA.

Lucko, G., Anderson-Cook, C. M., and Vorster, M. C. (2006). “Statistical considerations for predicting residual value of heavy equipment.” *Journal of Construction Engineering and Management*, 132(7), 723-732.

Lucko, G., Vorster, M. C., and Anderson-Cook, C. M. (2007). “Unknown element of owning costs—Impact of residual value.” *Journal of Construction Engineering and Management*, 133(1), 3-9.

Lucko, G., and Mitchell Jr., Z. W. (2010). “Quantitative research: Preparation of incongruous economic data sets for archival data analysis.” *Journal of Construction Engineering and Management*, 136(1), 49-57.

Lucko, G. (2011). “Modeling the residual market value of construction equipment under changed economic conditions.” *Journal of Construction Engineering and Management*, 137(10), 806-816.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). “Equation of state calculations by fast computing machines.” *The Journal of Chemical Physics*, 21(6), 1087-1092.

Mitchell Jr., Z. W. (1998). *A statistical analysis of construction equipment repair costs using field data and the cumulative cost model*, Doctoral Dissertation. Virginia Institute of Technology, Blacksburg, VA.

Ponnaluru, S. S., (2009). *Empirical analysis of used construction equipment and auction house revenues*, Doctoral Dissertation. Washington State University, Pullman, WA.

Ponnaluru, S. S., Marsh, T. L., and Brady, M. (2012). "Spatial price analysis of used construction equipment: the case of excavators." *Construction Management and Economics,* 30(11), 981-994.

R Core Team (2019). "R: A language and environment for statistical computing." <https://www.R-project.org/> (February 11, 2019).

Raiffa, H., and Schlaifer, R. (2000). *Applied statistical decision theory.* Wiley, Hoboken, NJ.

Robert C.P., and Casella G. (1999). *Monte Carlo Statistical Methods*. Springer, New York, NY

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

Vorster, M. C. (2009). *Construction equipment economics.* C.E.M.P Central Inc., Blacksburg VA.

Wang, L., and Zhang, X. (2018). "Bayesian analytics for estimating risk probability in PPP waste-to-energy projects." *Journal of Management in Engineering*, 34(6), 04018047.

Werner, M., Ji, W., and AbouRizk, S. (2018). “Improving tunneling simulation using Bayesian updating and hidden Markov chains.” *Proceedings of the 2018 Winter Simulation Conference*, IEEE, Piscataway, NJ, 3930-3940.

Zhang, L., Ekyalimpa, R., Hague, S., Werner, M., and AbouRizk, S. (2015). “Updating geological conditions using Bayes theorem and Markov chain.” *Proceedings of the 2015 Winter Simulation Conference,* IEEE, Piscataway, NJ, 3367-3378.

Zhang, S., Du, C., Sa, W., Wang, C., and Wang, G. (2013). “Bayesian-based hybrid simulation approach to project completion forecasting for underground construction.” *Journal of Construction Engineering and Management*, 140(1), 04013031.

Zhao, T., and Fu, C.C. (2006). “Infrastructure development and expansion under uncertainty: A risk-preference-based lattice approach.” *Journal of Construction Engineering and Management*, 132(6), 620-625.

Zong, Y. (2017). *Maintenance cost and residual value prediction of heavy construction equipment*, M.Sc. Thesis. University of Alberta, Edmonton, Canada.

3 CHAPTER 3: DATA-DRIVEN SIMULATION-BASED ANALYTICS FOR HEAVY EQUIPMENT LIFE-CYCLE COST

3.1 Introduction

Heavy-duty equipment plays a vital role in heavy civil construction and mining operations. Managing a heavy equipment fleet in a cost-efficient manner is key for long-term profitability. To ensure the cost-efficiency of equipment management, practitioners are required to accurately quantify the equipment life-cycle cost, instead of merely depending on the empirical method.

The profitability of a single job is, to a large extent, determined by decisions on equipment management. Appropriate decision-making is founded upon understanding the equipment costs and related impact factors (Halpin 2010). Construction practitioners who manage a large fleet of equipment are required to ensure long-term cost-efficiency in equipment utilization; however, practitioners often experience challenges in collecting and controlling equipment-related costs. Although equipment cost is normally over-simplified to be a unit rate in project bidding and management, it is essentially an aggregate of numerous small components such as fuel, oil, and lubricant costs (Vorster 2009). In practice, these scattered components are found in separate company systems, such as the equipment management system, purchase order system, or the accounting system. Consolidation of these scattered data can be complex and tedious, increasing the difficulty of cost-analysis and hindering efficient equipment management.

In equipment economics, it has been widely accepted that equipment costs can be broken down into two categories: ownership costs and maintenance costs, which may be termed slightly different (e.g. owning costs and operating costs), (Mitchell 1998). Studies on the ownership costs mainly focus on modeling the residual market value (RMV). RMV aims to estimate equipment value at any given equipment age (Lucko 2010; Vorster 2009). The transaction records from the auctioneers' databases have become the major data source for RMV studies, and advanced heuristic models have been developed for RMV predictions (Fan et al. 2008; Zong 2017).

On the other hand, the maintenance cost is more complex to unfold as its constituent parts are recorded internally in separate systems (i.e., equipment management system, accounting system, and etc.), which are the only data sources (Mitchell 1998). Traditionally, the cumulative cost model (CCM) is the primary tool to analyze the maintenance costs (Vorster 1980). Developed based on the CCM, the life-to-date (LTD) cost analysis can be applied when the practitioner has lifetime data. By contrast the period-cost-based (PCB) analysis can be applied only when short-term data are available (Mitchell et al., 2011; Bayzid et al. 2016).

Although efforts have been made to integrate the ownership cost model with maintenance cost model, two challenges exist: (1) the ownership and maintenance costs are simplified as continuous curves—which can generate a single cost value for any equipment age—but the variance (uncertainty) of the cost is overlooked in this model; and (2) the ownership and maintenance costs are obtained assuming the equipment is bought brand new and kept for usage until disposal; this is uncommon in practice. Therefore, the authors propose this data-driven simulation-based method with a perspective of equipment cost-analysis for the entire equipment life-span based on "used" conditions and considerations of uncertainties.

As a widely-accepted concept in construction engineering, the life-cycle cost analysis (LCCA) is to investigate total costs over the project life-span. Traditionally, in infrastructure projects, the analysis life-span could be a few decades with the consideration of construction and maintenance costs (Santos and Ferreira 2013). To examine the efficiencies of energy consumption for residential construction projects, LCCA can also be applied (Mithraratne and Vale 2004). Similarly, the concept of LCCA can be applied to analyze equipment cost. For example, the deterministic models can be used estimate the equipment life cycle cost (Mitchell 1998; Ghadam et al. 2012). One recent LCCA study aims to obtain the deterministic optimal life of equipment considering the uncertainty of inputs, such as fuel price and interest rate, but does not include the historical maintenance cost data (Gransberg and O′Connor 2015).

To fill the research gap and meet the aforementioned challenges, a data-driven simulation-based analytics approach is introduced to estimate the total cost at any point of equipment age incorporating ownership and maintenance cost, including variation (uncertainty). The K-means clustering and Expectation-Maximization (EM) algorithms are implemented to identify and quantify the turning points to distinguish different equipment maintenance stages. Monte Carlo simulation can be further applied to examine the equipment cost uncertainty. Through one numerical example, the proposed method is explained in detail. The functionality and validity of the proposed approach is implemented and tested in a practical case where the method is found capable of reflecting the life-cycle cost of equipment. In short, the proposed method can bring cost-related insight to decision-makers involved in equipment management.

3.2 Literature Review

In the literature review section, the authors intend to cover a variety of research areas that relate to this paper's content. This paper is interdisciplinary in nature in that integrates expertise from three main areas: (1) traditional equipment cost analysis; (2) simulation techniques in construction; and (3) life-cycle cost analysis. Although it is unambiguous that there are other areas of research that need to be covered in this section, they are respectfully excluded due to the length of the paper. The literature regarding to these three main areas are reviewed both separately and collectively below.

Equipment Ownership Cost

Equipment ownership costs normally consist of equipment depreciation (i.e. amortization) and fixed costs (Halpin 2010). The fixed costs—including interest, insurance, license and tax costs—can be calculated through the average value method, the average-annual value method, or the average market-value method. The calculated result values for each method are close enough to be virtually interchangeable (Vorster 2009). The fixed costs can also be assumed as a percentage (e.g. 13% to 18%) of the annual depreciation cost. The corresponding variance is determined by the interest rate, which reflects expectations on capital return (Halpin 2010). In short, the fixed costs vary depending on the construction market.

Depreciation is the difference between the purchase-price and salvage-value, which is the major driver for the equipment ownership cost (Mitchell 1998). However, equipment cost data stored in a construction company's system are limited, and a single mathematical model cannot be used to predict RMV for all heavy equipment (Fan et al. 2008). Fortunately, commercial software systems such as *EquipmentWatch* can easily collect transaction prices from auctions and private sales, drawn from multiple systems, and can consolidate them into one database when given an equipment model. The research on RMV analysis has been primarily conducted leveraging on the data from commercial systems (Lucko 2003). Taking advantage of these data, both mathematical and advanced heuristic models were proposed.

A multi-linear regression model for RMV prediction was firstly proposed based on the equipment age (Lucko 2003). A second-order polynomial RMV prediction model was proposed with the consideration of more factors—especially macro-economic indicators (Lucko et al. 2006). Among all the mathematical models, a popular one is the declining exponential model, which uses a single factor; here, the equipment age. The equipment age can be interpreted and represented by the service meter reading (Vorster 2009). Advanced models using heuristic algorithms were also developed, but such algorithms have not been widely used by practitioners. More recently, automated RMV prediction models based on two data-mining algorithms (autoregressive tree and random forest) have been developed, and have outperformed all the mathematical models (Fan et al. 2008; Zong 2017).

Both mathematical models and advanced heuristic models can estimate the residual value of equipment to provide analytical information for determining the ownership cost; however, one single value (rather than a value range) cannot interpret the uncertainties of the price. Purchasing and selling used equipment can be a capital-intensive process requiring the consideration of risk when creating prices. Construction companies are cautious about the variances of cost, i.e. the distributions of accumulated uncertainties.

Equipment Maintenance Cost

In equipment economics, the cost of equipment operations normally includes five components: (1) fuel, oil and lubricants; (2) tires or tracks; (3) ground engaging tools; (4) repairs; and (5) rebuilds (Mitchell 1998). Equipment maintenance cost typically refers to the operating cost excluding fuel, oil, and lubricants which account for the majority of operating costs (Peurifoy and Schexnayder 2002).

The cumulative cost model (CCM) was proposed to quantify and visualize the maintenance cost over the equipment age, but the CCM is not regulated as a specific mathematical model (Vorster 1980). A quadratic regression model was further developed based on the CCM (Mitchell 1998).

Furthermore, the life-to-date (LTD) analysis and period-cost-based (PCB) analysis are defined respectively based on the CCM (Mitchell et al. 2011; Bayzid 2014). The LTD approach requires the data of equipment for the whole usage life (Mitchell et al. 2011), while the PCB approach needs the data for any particular period of time (Bayzid 2014). Additionally, through CCM analysis, a trend is rising for cumulative maintenance cost over equipment age, as compared to the cumulative ownership cost with a declining trend. Combined together, the ownership and maintenance costs can yield a reasonable spot where the combined cost is lowest with a given equipment age, and often considered the optimal equipment age (Vorster 2009).

Compared to the ownership costs—especially RMV, which can be collected and shared by a third party—the maintenance costs are less likely to be shared among competitors; no mature third-party collects the cost information. To ensure both the quality and the interpretability, the model for maintenance cost is developed for specific types of equipment (model/make). The modeling of maintenance cost, therefore, must be conducted within the construction companies.

For instance, a total of 19 regression models have been developed for CCM analysis based on 270 machines (a 17-fleet combination) (Mitchell 1998). A recent study proposes a second-order polynomial regression model for cumulative maintenance cost of a CAT 785 based on 13-year span data collected from 26 pieces of the CAT 875 (Zong 2017).

Simulation Applications in Construction Engineering and Management

Simulation can provide computer-based representation of construction systems and processes to study the underlying behavior for decision support (Abourizk 2010). Simulation applications in construction engineering and management were originally inspired by the Monte Carlo simulation, which can repeat random sampling to obtain numerical results or probability distribution as outputs (Hastings 1970). With the rapid development of computer technology, the simulation application is developed with further complexity, examples of which can be found in the combination of various simulation types (e.g. discrete, continuous, hybrid, etc.), and other mathematical algorithms (e.g. optimization, data mining, and so on).

Specifically, the discrete event simulation (DES) approach is widely applied in achieving probability distributions and quantifying uncertainty through conceptualizing the logic of construction processes into object-oriented systems. Due to the nature of process modeling, DES has been predominant in providing decision support in the construction management domain. Resource-based simulation has shown great practical value in earthmoving operations (Hajjar and AbouRizk 1997). Simulations for tunnelling projects are have been developed to mimic the repetitiveness of operations inherent to construction processes (Er et al. 2000). Meanwhile, continuous simulation is also suitable for some construction activities, such as pavement construction (Puri et al. 2013). Although it is challenging to incorporate stochastic variability into continuous simulations, hybrid systems combining discrete and continuous simulations have been seen in earthmoving operations (Peña-Mora et al. 2008). Current challenges of construction simulation exist in the way of requiring significant efforts to establish simulation models and prepare model inputs. Therefore, the resilience and integration of simulation models has become important in more recent research endeavours (Abourizk 2010).

Life-Cycle Cost Analysis

During infrastructure construction, especially road and pavement construction, life-cycle cost analysis (LCCA) is a common practice for exploring cost-efficient alternatives. In 2006 in the United States, over 80% of all pavement projects from the state department of transportation conducted LCCA when selecting best practices (Chan et al. 2008). In these applications, the pavement costs are divided into construction costs and maintenance costs, similar to the division of ownership and maintenance costs for equipment. Based on this cost division, the pavement maintenance strategies, extended pavement preservation strategies, and rehabilitation strategies were compared to make project decisions (Mandapaka et al. 2012). A specific example is comparing perpetual pavement (thick pavement used at the construction stage) to conventional pavement used to maintain roads (Amini 2011).

In residential construction, the environmental impact of residential houses in New Zealand was deemed as “costly” and studied through life-cycle analysis (Mithraratne and Vale 2004). This shows a wide range of LCCA applications in the construction industry, which can be also introduced into equipment cost analysis. A non-liner regression model is established to achieve deterministic life cycle cost (Ghadam et al. 2012). Since the historical maintenance data is limited, a recent LCCA study exploits other cost data, such as fuel price and tire cost, to estimate the maintenance cost (Gransberg and O′Connor 2015). By integrating stochastic method with deterministic model, this study achieves the optimal equipment life by taking the uncertainties of maintenance cost inputs into account. However, to obtain the range of equipment life-cycle cost instead of a deterministic value, one major challenge is to utilize both ownership and maintenance cost.

In summary, maintenance and ownership costs were mostly studied and modelled separately in the previous research; however, further investigation of combining both cost models in order to develop a mathematical model remains unsolved. The literature has shown that the level of difficulty to obtain the maintenance data has prevented researchers from developing models incorporating both costs. In addition, to simply represent the incurred equipment costs with a single or a static set of values, interpreted directly from the cost model, can be biased. The uncertainty (risk) enclosed in the equipment cost is hard to obtain and, consequently, affects decision making. Using the outline presented herein of the existing challenges and research gap, the authors propose simulation-based analytics to quantify the uncertainty associated with the equipment cost; and to represent cost with a floating range rather than with limited static values.

3.3 Methodology

A data-driven, simulation-based analytics is proposed to quantify the cumulative cost of equipment associated with uncertainties. The workflow of the methodology is summarized in Figure 3.1.

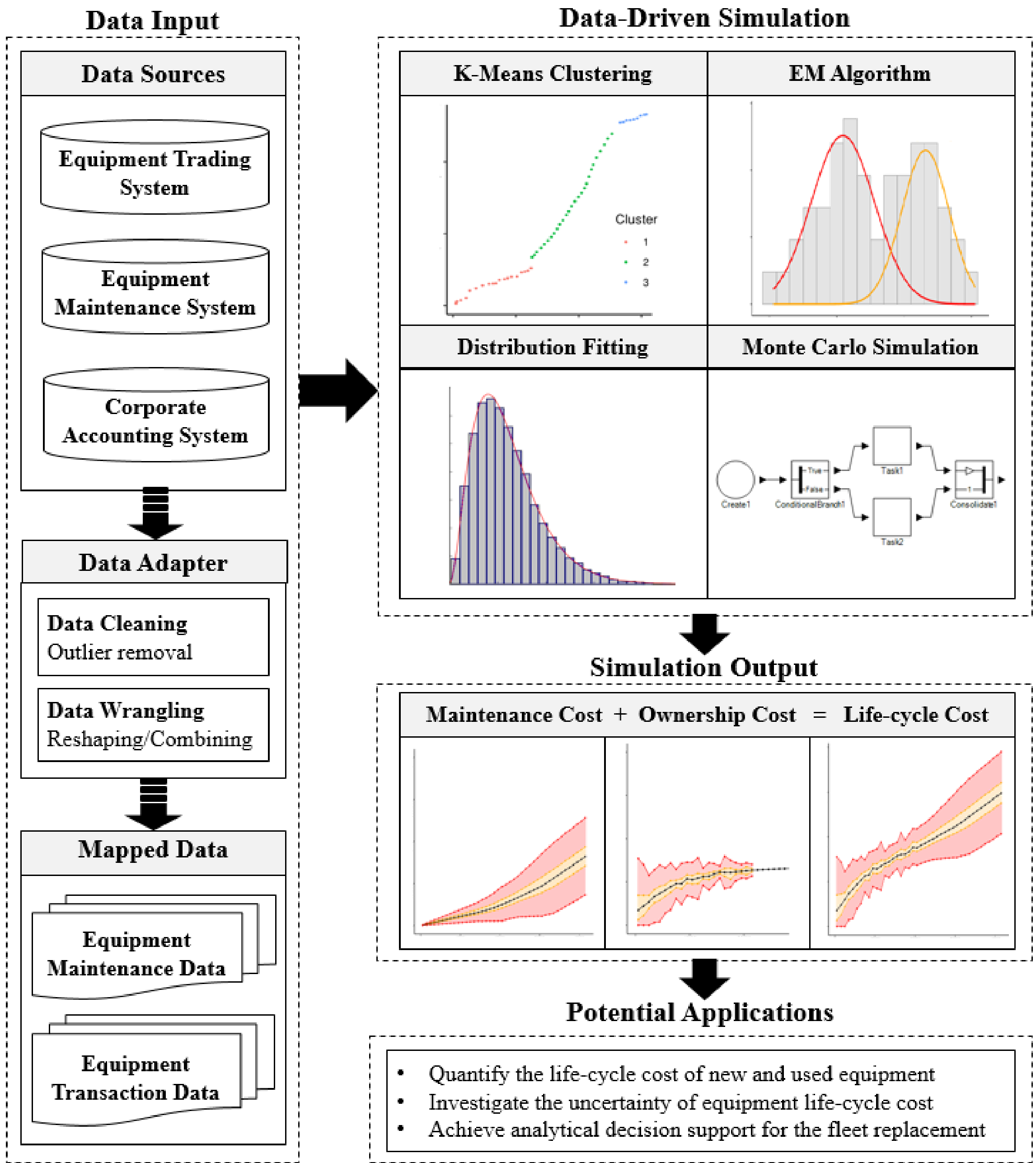


Figure 3.1 Methodology Flowchart of Data-Driven Measurements Method

The ownership and maintenance cost data are collected from an equipment trading system, an equipment maintenance system, and a corporate accounting system, respectively. After data cleaning and wrangling, maintenance cost data for each piece of equipment are clustered into several stages using the K-means clustering method. The Expectation-Maximization (EM) algorithm is applied to generate the distributions of turning points among stages.

A numerical example will be presented to explain the implementation of the K-means clustering and EM algorithm step-by-step. At the same time, the ownership cost data are grouped based on specific service meter rating (SMR) intervals (e.g. 500 hour) and the data within the same interval are fitted into distributions as the simulation inputs. Lastly, the Monte Carlo simulation model is applied. Through multiple iterations, results are generated to evaluate the life-cycle cost of equipment by combining both maintenance and ownership costs. In this study, the implementation of the K-means clustering, EM algorithm, and simulation are conducted using the *R* programming language, an open-sourced software for statistical computing and graphics (*R* Core Team 2019). Noted that algorithms are generic, and can be achieved and implemented using other types of coding and data-mining programs.

Data Input

The inputs, maintenance cost data and ownership cost data, are collected from various sources. The ownership cost (depreciation in this study) is determined by the purchase price and salvage value of equipment as Eq. 3.1, which was formulated similarly to the equipment amortization (Mitchell 1998).

$$OC_t = P_0 - S_t \tag{3.1}$$

where OC_t is the ownership cost, namely depreciation cost, at time t. P_0 is the initial purchase price and S_t is the salvage value of the machine at time t.

The purchase price is often recorded in the company's accounting system. Salvage value of equipment can be collected from three data sources. The first data source, in practice, is the accounting system as internal data are the most readily available (Mitchell 1998; Zong 2017). As a note, after much data handling, the data stored in the accounting system may be changed or even deleted by accident, which can lead to inconsistency (Mitchell 1998). The second data source of equipment salvage values is collected by equipment auctioneers; hundreds of pieces of equipment are traded with accompanying information recorded in the auctioneers' databases. Furthermore, the auctioneer data are consolidated into large databases by commercial third-party companies. With substantial subscription fees, researchers can access data that are comprehensive and ready for use. Particularly, in the aforementioned situation of data scarcity, a third-party data provider can be a decent alternative for equipment data analysis; the authors recommend using third-party providers due to their overall data integrity.

Data cleaning and wrangling may be necessary depending on the quality and integrity of the collected data. In this study, the following four factors are required as the inputs for the ownership cost: transaction date, the manufacturer/model, service meter reading (SMR), and the selling price of the equipment. A sample dataset is provided in Table 3.1 as an illustrative example of data required for the analysis.

Table 3.1 Sample Input Data of Equipment Transaction

Date	Manufacturer/Model	SMR (hour)	Price (CAD)
09-Mar-16	Hitachi EX1200	3658	$82,000.00
16-Oct-17	Hitachi EX1200	8695	$46,000.00
06-May-18	Hitachi EX1200	12326	$57,500.00

The equipment maintenance cost, which is often not open to the public and is kept within the company, can only be collected internally from the construction companies (Bayzid 2014; Zong 2017). Companies may not always store all the maintenance data in the same database, and some data wrangling is therefore normally required (Mitchell 1998). The service meter reading (SMR), commonly recorded in different sources, is critical for data wrangling. Five factors are required as the inputs for the maintenance cost to implement the proposed method, namely the equipment identification, the manufacturer/model, SMR, the date of maintenance activity, and the maintenance cost as shown in the sample in Table 3.2.

Table 3.2 Sample Input Data of Equipment Maintenance

Equipment ID	Manufacturer/Model	SMR (hour)	Date	Price (CAD)
A	Hitachi EX1200	1053	01-Dec-15	$6,907.45
A	Hitachi EX1200	2315	01-Jan-16	$4,860.65
A	Hitachi EX1200	3503	01-Feb-16	$14,405.77

Data-Driven Simulation

K-means Clustering

After obtaining equipment maintenance cost data, the cumulative maintenance cost of each individual piece of equipment is investigated. For each piece, the cumulative maintenance cost increases with equipment age, typically represented by the equipment service meter reading (SMR) in units of hours or kilometers. In this study, the K-means clustering algorithm is applied to divide the cumulative maintenance cost into stages and to find defining turning points among the stages (see Fig. 3.3 for details). The *stage* is defined as a SMR range, within which the slope of cost increase remains stable. At the *turning point*, a sudden cost rise occurs due to major repairs, such as rebuilds.

In data mining, clustering refers to a broad set of techniques for finding subgroups (clusters) in a given dataset. Clustering leads to the discovery of previously unknown groups within data (Ji et al. 2018). Traditionally, clustering methods can be categorized into either a partitioning-based clustering method or a density-based clustering method (Han et al. 2011). K-means clustering, the partitioning-based clustering method, aims to segregate observations into a pre-specified number of clusters, namely K (Forgy 1965). The essential idea of K-means clustering is that a reasonable cluster is within the minimal within-cluster variance (WCV) as illustrated in Eq. 3.2, assuming that the Euclidean distance is used for measuring the distances between observations. When conducting K-means clustering, each observation is randomly assigned a number, from 1 to K and, through iterations, clustering re-assignment continues until the minimal *WCV* is achieved. The objective function, defined by K-means clustering, is illustrated in Eq. 3.3.

$$WCV(C_k) = \frac{1}{|C_k|} \cdot 2 \sum_{i \in C_k} \sum_{j=1}^{P} (x_{ij} - \bar{x}_{ij})^2 \qquad (3.2)$$

$$\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} \{WCV(C_k)\} \qquad (3.3)$$

where $|C_k|$ is the number of observations in the k^{th} cluster, i is the i^{th} observation in the cluster C_k, j is the j^{th} feature, P is the total number of features, x_{ij} is the value of feature j for the observation i, $\bar{x}_{ij}$ is the mean for feature j in cluster C_k, and K is the pre-specified clusters.

In practice, it is reasonable to divide the maintenance of heavy civil equipment into stages, which are normally divided by certain events. During the early equipment lifespan, it can be foreseen that maintenance costs after the manufacturer warranty period will greatly increase. Similarly, rebuilds, namely replacing the engine or other major components, may significantly increase cost (Mitchell 1998). According to the repair limit theory, there is a limit on the amount of money which can be spent on the repair of a vehicle at any particular job and therefore the maintenance stages are limited (Drinkwater and Hastings 1967).

In generic cases, the number of clusters, namely the number of maintenance stages, is dependent on and defined by collected historical maintenance data. As such, the *K* value is considered from 2 to 10. To be specific, 10 is the maximum number of stages assumed for equipment maintenance, which may be adjusted under certain circumstances. The objective function (to find the optimal number of clusters) aims to maximize the average of all costs (combining ownership and maintenance costs) incurred at the turning points, illustrated as Eq. 3.4.

$$\max_{K=2,\dots,10} \bar{C} = \max_{K=2,\dots,10} \frac{1}{(K-1)} \cdot \sum_{k=1}^{K-1} \left| \min\left(CMC_{C_{k+1}}\right) - \max\left(CMC_{C_k}\right) \right| \quad (3.4)$$

where $\bar{C}$ is the average cost incurred at the turning points, K is the number of clusters and CMC_{C_k} is the cumulative maintenance cost (*CMC*) of observation in the k^{th} cluster.

Expectation-Maximization Algorithm

Through the K-means clustering, the optimal number of clusters and the turning points between clusters for each individual piece of equipment can be determined. The SMR of the turning points can indicate the age when the rebuild is typically made, the age when the maintenance costs have significantly increased, or can provide analytical decision support for equipment maintenance strategies. For example, it is economical to sell a piece of equipment at a particular SMR (before the major rebuild). After combining the service meter readings (SMRs) of turning points for many pieces of equipment of the same type, a distribution can be generated with two or three average turning points. Rebuild can be conducted twice or three times in general for a typical piece of heavy civil equipment before disposal, such as a mining truck. This distribution is a typical mixture model, which is a mixture of two or more probability distributions.

The Expectation-Maximization (EM) algorithm can be used to separate a mixture model into components in terms of its distributions. In this study, the EM algorithm is applied to a one-dimensional dataset: the service meter readings at turning points. Given an observed dataset, *Z*, the EM algorithm aims to find the maximum likelihood estimate of probabilistic models, following expectation and maximization steps (Dempster 1977). Maximum likelihood estimate (MLE) of unknown parameters is determined by maximizing the marginal likelihood of the observed data *Z*:

$$\max_{\theta} \log \int_{x} P(x, z|\theta) dx \tag{3.5}$$

where θ is the parameters of the probabilistic model, *x* is unobserved variables and *z* is observed variables from dataset *Z*.

Based on the Eq. 3.5, the EM algorithm finds the MLE of the marginal likelihood through iteratively applying expectation and maximization steps as below:

1. Expectation step: Compute $Q(x) = P(x|z, \theta)$
2. Maximization step: Compute $\theta = \text{argmax}_{\theta} \int_{x} Q(x) \log P(x, z|\theta) dx$

Distribution Fitting

For the input modeling stage of simulation, a probability distribution is usually fitted based on the data available (Leemis 2004; Kuhl et al. 2009). There are 30 to 40 standard probability distributions available for distribution fitting in the commercial software (Biller and Gunes 2010). The aim of input modeling is to estimate the parameters of the selected probability distributions. Common methods of input modeling include: (1) the maximum likelihood estimation method; (2) the matching moment method; (3) the matching percentiles method; and (4) the least squares estimation method (Law 2007). After obtaining the parameters, the common methods of assessing the goodness-of-fit for a probability distribution include Kolmogorov-Smirnov and Anderson-Darling tests. The distribution can also be tested through visualization and density-histogram plots, or a probability plot (Biller and Gunes 2010). The goodness-of-fit tests determine the level of confidence for the generated distribution fit for rejection or acceptance.

In this study, the ownership cost data are divided into groups based on the same SMR interval; in this case, every 500 engine hours is a reasonable interval suggested by the past research (Mitchell 1998). In each interval group, the residual value percentage is fitted into a probability distribution. The parameters of distribution are estimated through the maximum-likelihood estimate method; goodness-of-fit is tested using the Kolmogorov-Smirnov test.

Monte Carlo Simulation

Monte Carlo simulation can repeat random sampling to obtain numerical results or probability distributions (Hastings 1970). In our case, the cumulative maintenance and cumulative ownership costs are the intended outputs from the simulation. The cumulative ownership cost is illustrated in the Figure 3.2. According to the Eq. 3.1, ownership cost at equipment age t is the difference between the purchase price (P_0) and the salvage value (S_t). The original purchase price is fixed and known, and the salvage value is unknown. Through distribution fitting, the probability distribution of salvage value can be obtained given the available equipment transaction data. For example, in each simulation iteration, the salvage value for equipment at age SMR_t, namely S_{SMRt}, is a random number sampling with a confident fitting from a Beta distribution. Thus, the cumulative ownership cost is the difference between a fixed number for purchase value and a random number for salvage value (i.e. $MC_{SMRt}=P_0-S_{SMRt}$).

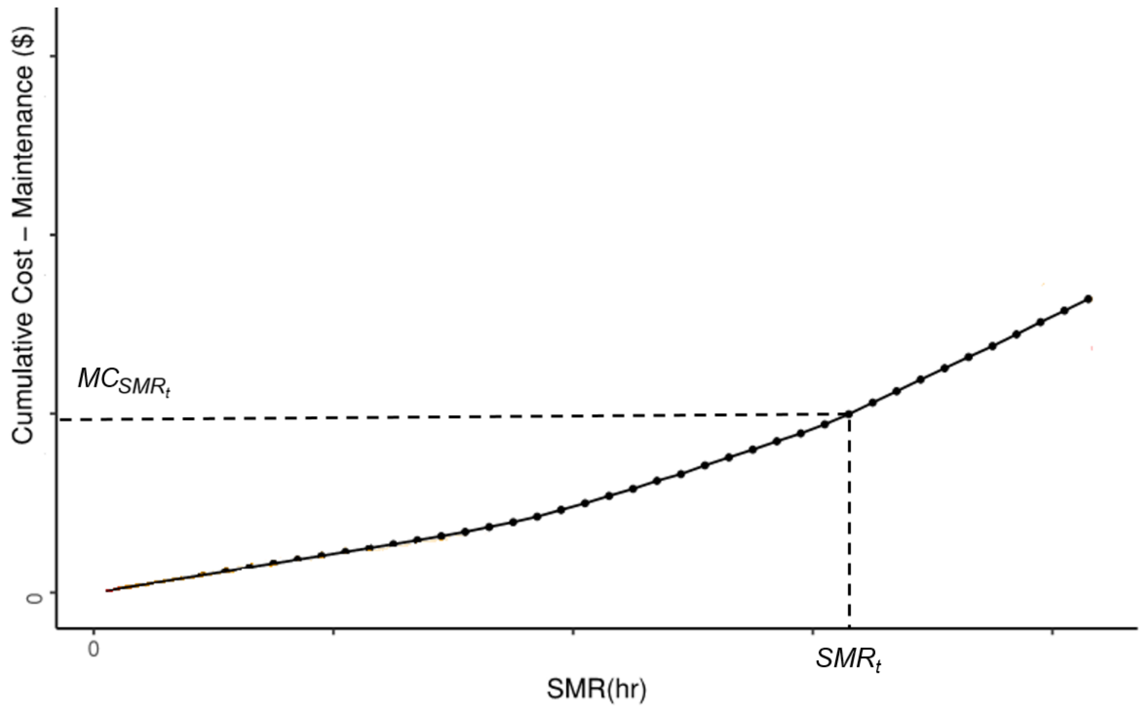


Figure 3.2 Illustration of Simulation for the Cumulative Ownership Cost

Additionally, the Monte Carlo simulation of maintenance cost is explained in steps as demonstrated in Figure 3.3, with an assumption that the maintenance of a piece of equipment contains three stages separated by two major rebuilds. In the first step, turning points between stages in hours, namely $SMR_{S1\text{-}S2}$ and $SMR_{S2\text{-}S3}$, are randomly sampled from two probability distributions through the K-means clustering and EM algorithms. In the second step, costs incurred at the turning points, $C_{S1\text{-}S2}$ and $C_{S2\text{-}S3}$, are achieved through random sampling from probability distributions fitting the maintenance cost data. In the third step, the increase rate of maintenance cost (S_{S1}, S_{S2}, and S_{S3}), which is the slope of line within each stage, is also generated as a randomly sampled value from probability distributions based on the historical data. In each iteration of the simulation, the cumulative maintenance cost at any equipment age is calculated using the aforementioned variables obtained through random sampling. As such, the cumulative cost of maintenance (MC_{SMRt}) is calculated for any equipment SMR (SMR_t), incorporating the probabilities of each variable, as shown in Fig. 3.3.

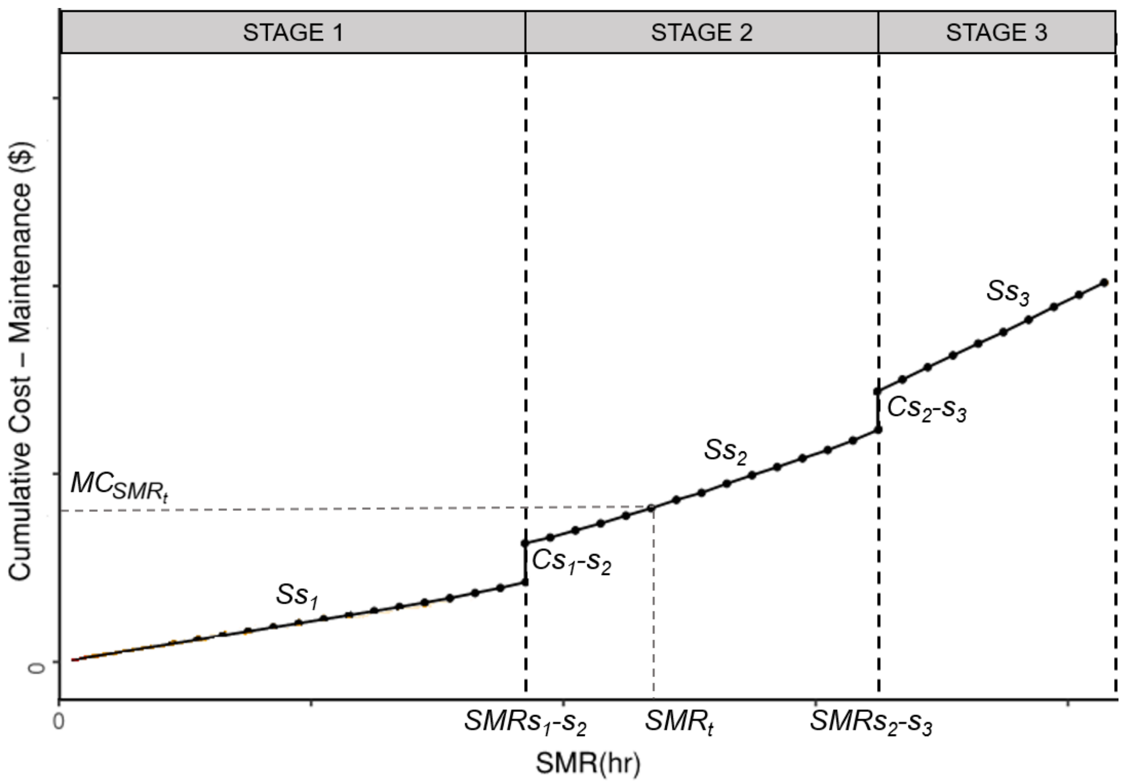


Figure 3.3 Illustration of Simulation for the Cumulative Maintenance Cost

Data Output

Based on the Monte Carlo simulations, cumulative maintenance cost and cumulative ownership cost at any equipment age (SMR_t) is obtained. As such, the cumulative total cost (TC_{SMRt}) can be achieved as the sum of maintenance cost (MC_{SMRt}) and ownership cost (OC_{SMRt}) through Eq. 3.6.

$$TC_{SMRt} = OC_{SMRt} + MC_{SMRt} \tag{3.6}$$

Through a reasonable number of iterations, the probability distribution of cumulative maintenance and ownership costs are developed respectively, and the probability distribution of the cumulative total costs are obtained as the inputs and outputs of the simulation, separately.

3.4 Illustrative Example

K-means Clustering

This example intends to illustrate how the proposed methodology and algorithms are applied. To start with, the historical maintenance cost for one piece of heavy equipment is listed as the first 2 columns of Table 3.3. Then the K-means clustering algorithm is applied, and the results based on different *K* values are summarized in last 3 columns in Table 3.3 (K= 2, 3 or 4). Theoretically, the value *K* can vary from the past and available data. These results are also illustrated as scatterplots marked in different colors in Figure 3.4, 3.5 and 3.6. *K* less than 5 represents the clusters well in the example according to the Figure 3.4, 3.5 and 3.6.

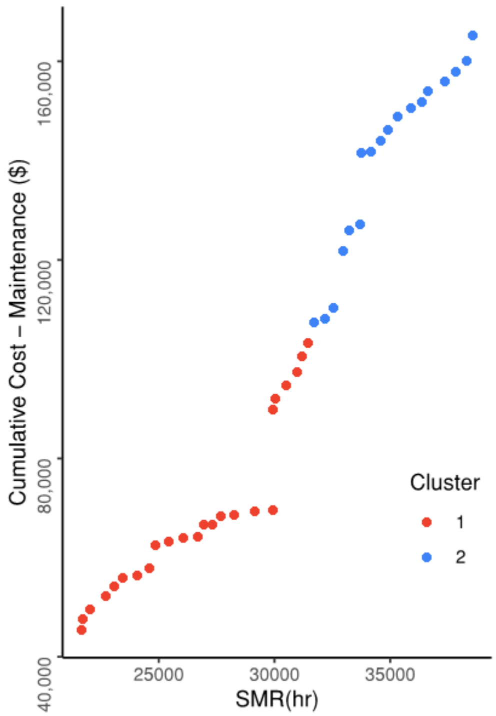


Figure 3.4 Scatter Plot Based on K-means Clustering Results (K=2)

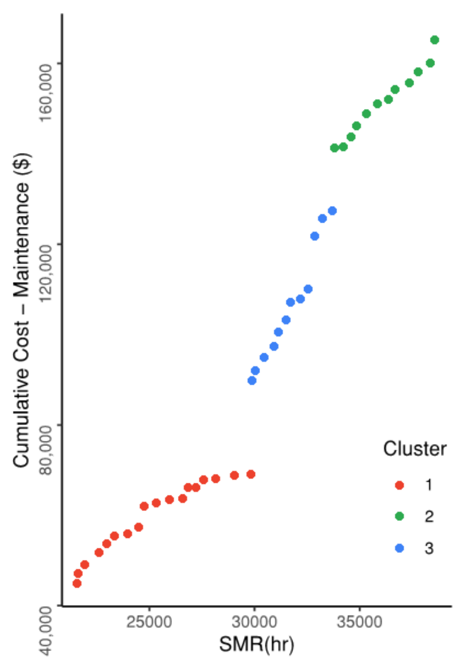


Figure 3.5 Scatter Plot Based on K-means Clustering Results (K=3)

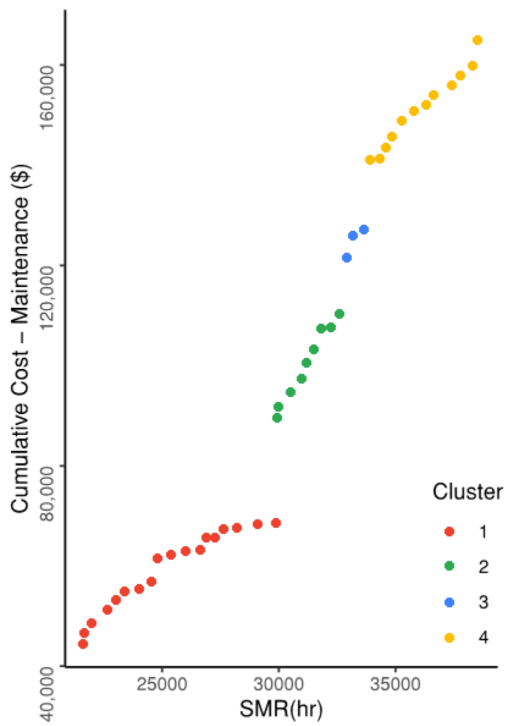


Figure 3.6 Scatter Plot Based on K-means Clustering Results (K=4)

Table 3.3 Maintenance Cost Data of One Machine

SMR (hr)	Cumulative Maintenance Cost (CAD)		Cluster (K=2)	Cluster (K=3)	Cluster (K=4)
21692	$	40,317	1	1	1
21719	$	41,376	1	1	1
21719	$	42,085	1	1	1
22006	$	43,316	1	1	1
22694	$	46,119	1	1	1
23152	$	48,196	1	1	1
23389	$	49,035	1	1	1
24077	$	49,462	1	1	1
24593	$	50,816	1	1	1
24897	$	55,066	1	1	1
25452	$	55,751	1	1	1
26038	$	56,223	1	1	1
26652	$	56,590	1	1	1
27003	$	58,719	1	1	1

SMR (hr)	Cumulative Maintenance Cost (CAD)		Cluster (K=2)	Cluster (K=3)	Cluster (K=4)
27331	$	58,936	1	1	1
27669	$	60,123	1	1	1
28261	$	60,340	1	1	1
29126	$	60,834	1	1	1
29883	$	61,062	1	1	1
29883	$	79,188	1	2	2
29886	$	79,381	1	2	2
29949	$	80,373	1	2	2
30449	$	83,412	1	2	2
30983	$	85,777	1	2	2
31153	$	88,440	1	2	2
31488	$	90,730	1	2	2
31824	$	94,568	2	2	2
32034	$	94,777	2	2	2
32546	$	96,950	2	2	2
32878	$	107,220	2	2	3
32970	$	107,226	2	2	3
33202	$	110,885	2	2	3
33663	$	111,716	2	2	3
33826	$	124,608	2	3	4
34204	$	124,622	2	3	4
34626	$	126,519	2	3	4
34832	$	128,555	2	3	4
35313	$	131,049	2	3	4
35851	$	132,965	2	3	4
36355	$	133,826	2	3	4
36621	$	135,560	2	3	4
37379	$	137,146	2	3	4
37763	$	139,001	2	3	4
38338	$	140,788	2	3	4
38541	$	145,264	2	3	4

According the objective function as illustrated in Eq. 3.4, the average of costs incurred at the turning points are calculated for three clustering results using Eq. 3.7 to 3.9. When conducting K-means clustering with 3 clusters (i.e. K=3), the average cost ($\bar{C}$) based on the turning points are calculated as a result of $15,509, which is the largest. Following the Eq. 3.4, the equipment maintenance should be divided as three stages.

$$\bar{C}_{K=2} = \frac{1}{(K-1)} \cdot \sum_{k=1}^{K-1} \left| \min\left(CMC_{C_{k+1}}\right) - \max\left(CMC_{C_k}\right) \right| = |94{,}568 - 90{,}730| = \$3{,}838 \tag{3.7}$$

$$\bar{C}_{K=3} = \frac{1}{(3-1)} (|79{,}188 - 61{,}062\,| + |124{,}608 - 111{,}716|) = \$15{,}509 \tag{3.8}$$

$$\bar{C}_{K=4} = \frac{1}{(4-1)} (|79{,}188 - 61{,}062\,| + |107{,}220 - 96{,}950| + |124{,}608\, - 111{,}716|) = \$13{,}763 \tag{3.9}$$

Expectation-Maximization Algorithm

In the available historical data, there are a total of 23 pieces of the same equipment (make/model), and the maintenance data of these pieces of equipment are available. The turning points between stages are calculated for each machine through K-means clustering method and the corresponding results are listed in the Table 3.4. For example, for the Machine 3, there are 4 stages with three turning points throughout all 4 stages; Machine 6 has only one turning point between the two stages. In total, there are 49 turning points as listed in Table 3.4, for 23 machines respectively.

Table 3.4 Turning Points of 23 Machines

Machine ID	$SMR_{S1\text{-}S2}$ (hr)	$SMR_{S2\text{-}S3}$ (hr)	$SMR_{S3\text{-}S4}$ (hr)
1	22256	32571	38807
2	29883	33663	-
3	27265	31489	37715
4	29778	31289	-
5	22155	33595	40573
6	23481	-	-

Machine ID	$SMR_{S1\text{-}S2}$ (hr)	$SMR_{S2\text{-}S3}$ (hr)	$SMR_{S3\text{-}S4}$ (hr)
7	30966	36426	-
8	28031	33754	-
9	20779	26331	32221
10	28721	33441	36572
11	27943	32616	35121
12	29358	34928	-
13	18763	-	-
14	10590	17001	20970
15	21627	-	-
16	16909	-	-
17	21678	-	-
18	15893	22373	-
19	20606	-	-
20	13800	20414	-
21	16520	22254	-
22	18494	21624	25331
23	14265	20933	24569

The service meter readings of 49 turning points, as summarized in Table 3.4, are then plotted as a histogram showing the mixture model, which consists of two normal distributions. The EM algorithm is applied to find the maximum-likelihood estimate of the probabilistic models. As a result of the EM algorithm, two normal distribution components are obtained from the plotted histogram to represent the SMR of turning-point occurrence in general, shown in Figure 3.7. For this type of equipment, two normal distributions can represent the maintenance stages and related SMRs (as distributions) when major rebuilds are expected. By repeating this calculation, other types of equipment can be studied. In the next section, a more detailed case will be presented to show how to eventually make use of the simulation results.

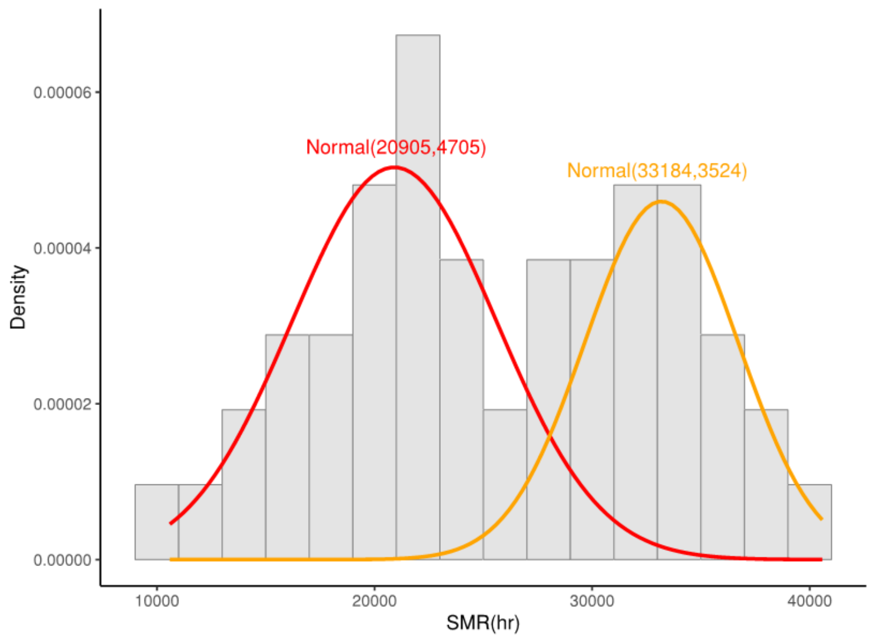


Figure 3.7 Results of the EM Algorithm

3.5 Case Study

In this case study, the maintenance cost data for 28 pieces of mining trucks of the same model and make were collected from a construction contractor in Alberta, Canada. The maintenance cost data has been extracted internally from the equipment management and accounting systems. The ownership cost data, namely equipment transaction records, were extracted from a third-party commercial system subscribed to by the same company. The cost data from different systems were then cleaned, combined, and rid of outliers (e.g. missing data). In the analysis, the annual interest rate is assumed to be 3% to incorporate the impact of time on price. The original cost data in this case study (in CAD) has been normalized (i.e. divided by a constant) due to the confidentiality requirement.

Following the proposed methodology, K-means clustering is firstly applied to each piece of equipment, based on the maintenance cost data, to extract the turning points among the maintenance stages. After combining all the service meter readings of turning points in hours for the whole fleet, the mixture model (namely Gaussian mixture model) is formed and the EM algorithm is used to estimate the parameters of two normal distribution components. The other parameters (Figure 3.3) are achieved as simulation inputs through probability distribution fitting. Based on the maintenance cost data collected for 28 machines, the distribution of inputs for random sampling are achieved and summarized in Table 3.5. For the ownership cost, the residual value percentage of equipment is sliced into 80 groups based on equal-sized SMR intervals of 500 hours per interval. For the data in each group, probability distribution fitting is applied to estimate the parameters and to check the goodness-of-fit. 1000-iterations are conducted for the simulation, and results for maintenance and ownership costs are shown in Figures 3.8 and 3.9.

Table 3.5 Distribution of Input Modeling

Variable	Description	Unit	Distribution
$SMR_{S1\text{-}S2}$	Turning point between Stages 1 and 2	Hours	Normal(20920,3530)
$SMR_{S2\text{-}S3}$	Turning point between Stages 2 and 3	Hours	Normal(33195,4742)
$C_{S1\text{-}S2}$	Cost occurred at $SMR_{S1\text{-}S2}$	CAD	Gamma(2.13,102956)
$C_{S2\text{-}S3}$	Cost occurred at $SMR_{S2\text{-}S3}$	CAD	Gamma(1.97,114843)
S_{S1}	Cost increasing rate for Stage 1	CAD/hour	Gamma(1.28,19.67)+53.34
S_{S2}	Cost increasing rate for Stage 2	CAD/hour	Gamma(1.00,85.47)+56.19
S_{S3}	Cost increasing rate for Stage 3	CAD/hour	Gamma(1.05,32.03)+179.95

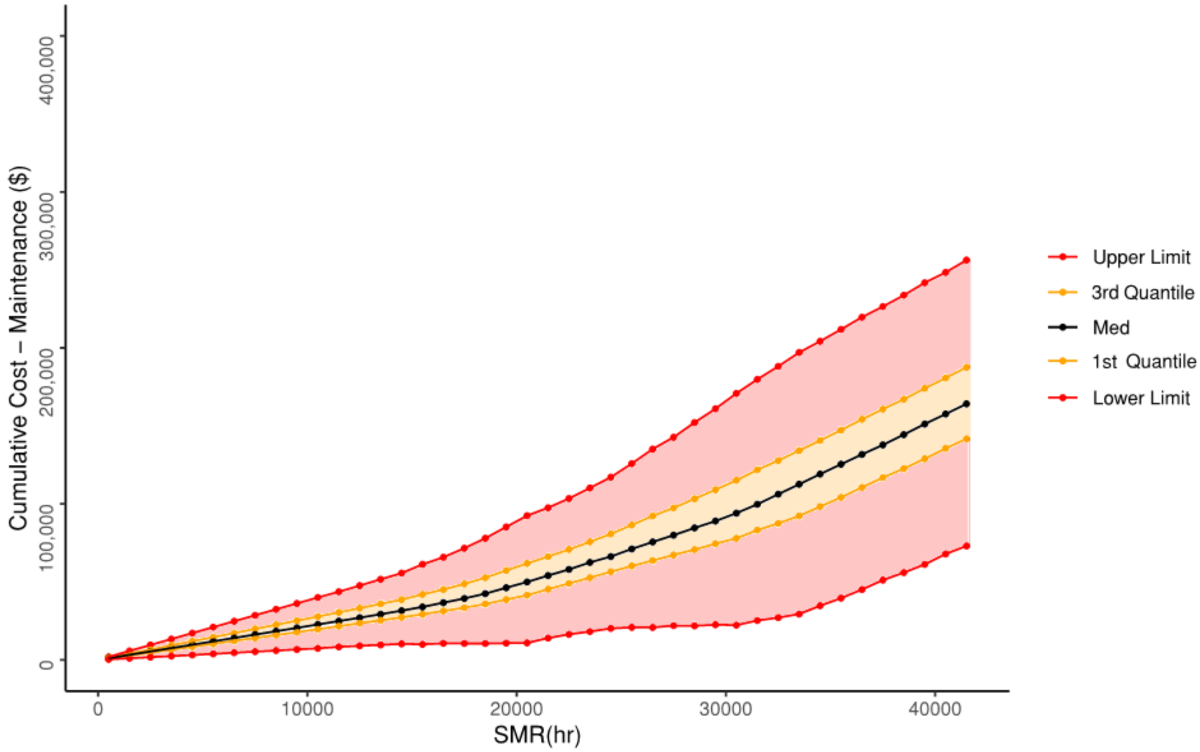


Figure 3.8 Simulation Results of Cumulative Maintenace Cost

For any given time point of engine hour (SMR_t), there are 1,000 simulated values for the cumulative maintenace cost. The distribution of 1,000 simulated values can be displayed as a box plot summarizing the: (1) maximum (upper limit); (2) third quantile; (3) median; (4) first quantile; and (5) minimum (lower limit). In Figure 3.8, the dotted line marked in black is the median of the cumulative maintenance costs. The dotted line marked in orange above the median is the 3rd quantile and the other in orange below the median is 1st quantile. The upper and lower limits are marked in red. Due to use conditions and maintenance strategies, for different pieces of the same equipment the cumulative maintenace cost is likely to fall into the orange area between 3rd and 1st quantiles. There is still a chance, however, that cumulative maintenance costs will fall into the red area, above 3rd quantile or below 1st quantile.

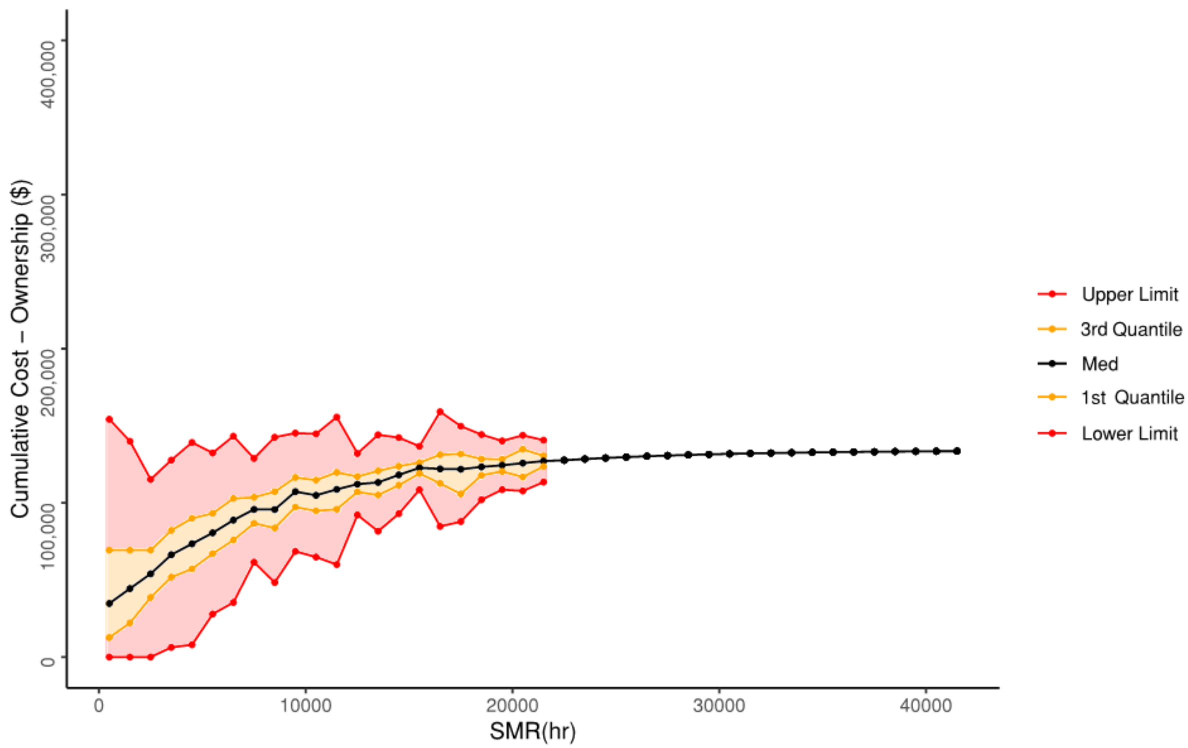


Figure 3.9 Simulation Results of Cumulative Ownership Cost

For the cumulative ownership cost, similarly, the simulation results are illustrated in Figure 3.9 where red, orange, and black have the same meanings as they have in Figure 3.8. Notably, after about 22,000 equipment service hours, residual market values become rare and therefore it is assumed to become steady.

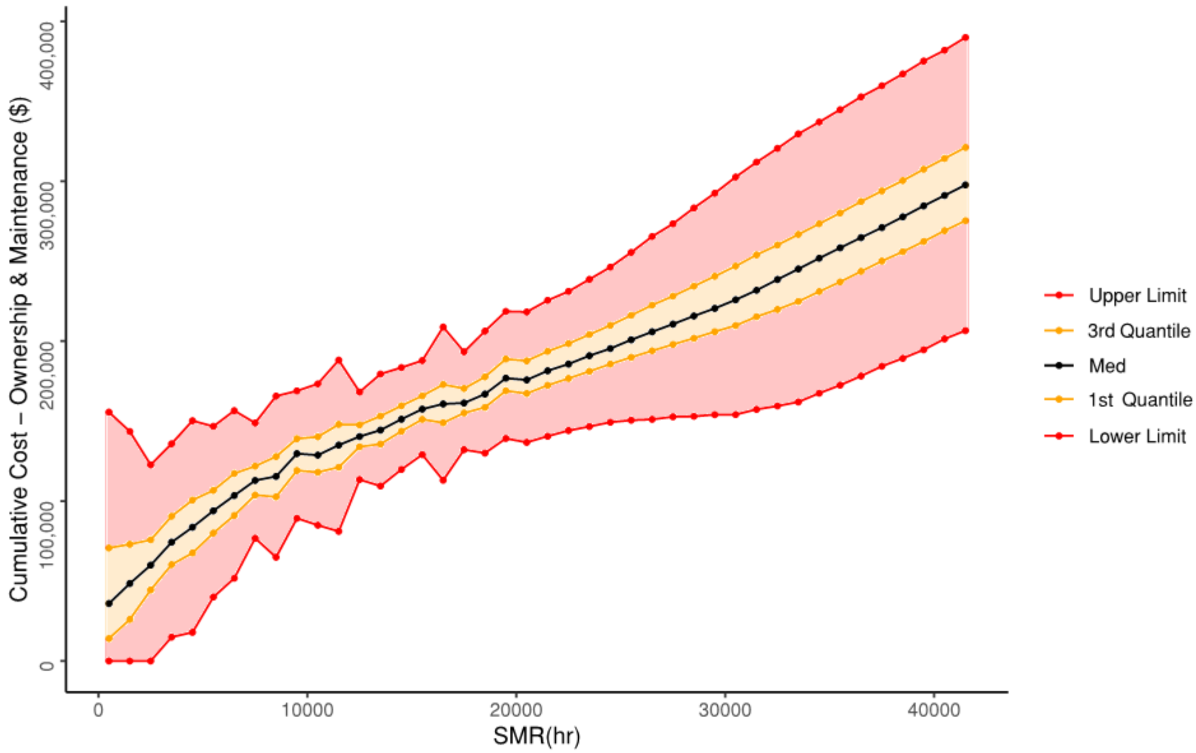


Figure 3.10 Simulation Results of Cumulative Total Cost

According to the Eq. 3.3, the cumulative total cost can be calculated as the sum of cumulative maintance costs and cumulative ownership costs. After combing the values of Figures 3.8 and 3.9, Figure 3.10 is generated to represent the cumulative total cost: the dotted line marked in black is the median of total costs while the dotted lines marked in orange represent the 3rd quantile and 1st quantile, respectively. In addition, the dotted lines marked in red are the upper limit and lower limit of cumulative ownership costs, respectively. It can be observed that the interquartile range (IQR) between the 3rd and 1st quantiles for the equipment between the 10,000 and 20,000 hr is remarkably smaller than the equipment before 10,000 hr and after 20,000 hr; this means that the cost of ownership and maintenance is more predictable in the middle of the equipment life-span than at earlier and later time frames of equipment usage.

Discussions

In equipment management, a dilemma often exists in equipment purchase: When is the optimal time for purchase? The life-cycle cost (i.e. cumulative total cost) of used equipment and new equipment can be further compared using this data-driven, simulation-based analytics to provide insight into the optimal purchase time.

Say there are five simulation scenarios based on the case study discussed herein: (1) the cost of a brand new equipment purchased at 0 engine hours is \$160,000; (2) the cost of a used equipment purchased at 1,000 engine hours is \$118,000; (3) the cost of a used equipment purchased at 5,000 engine hours is \$82,000; (4) the cost of a used equipment purchased at 10,000 engine hours is \$59,000; and finally (5) the cost of a used equipment purchased at 15,000 engine hours is \$41,600. Note that the purchase cost of brand-new equipment matches the listing price while the purchase cost of used equipment is based on the median cost of similar used equipment among historical equipment transaction records.

As illustrated in Figure 3.11, the median cumulative costs over the equipment service reading can be plotted for these five scenarios with different colors. In this case, the service meter reading is taken relative to the point in time when the equipment is purchased; thus, it starts from 0 hours for both brand new and used equipment. In other words, even for the used equipment purchased at 15,000 engine hours, the cumulative cost grows relatively after the 15,000 engine hours.

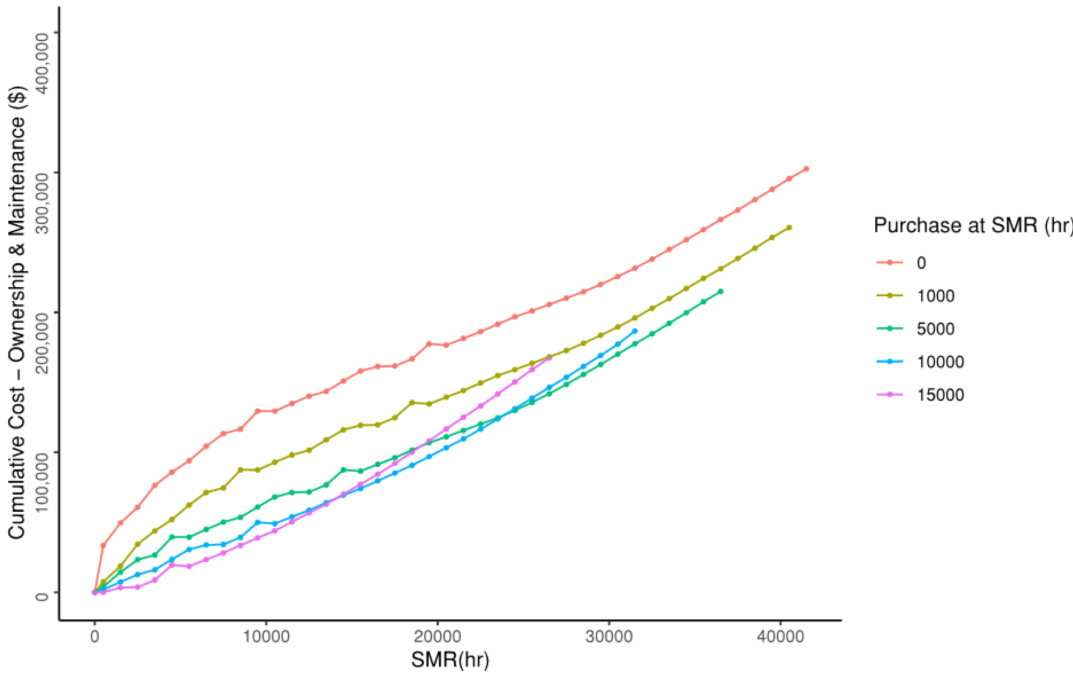


Figure 3.11 Cumulative Cost for Equipment Purchased at Various SMR

In Figure 3.11, the interquartile range (IQR) is not shown so as to keep the graph less crowded, and more straightforward to understand. Figure 3.12 explicitly shows the IQR and the dashed lines illustrate the average value for each of the individual scenarios. The expected usages after purchasing the equipment are 5,000 hours, 15,000 hours and 25,000 hours, which are respectively marked in red, orange, and yellow. According to the boxplots of the cumulative cost in Figure 3.12, if the expected usage is only 5,000 hours, purchasing the used equipment at 15,000 engine hours is most cost-efficient. On the other hand, if the expected usage is 25,000 hours, purchasing the used euqipment at 10,000 hours is more cost-efficient.

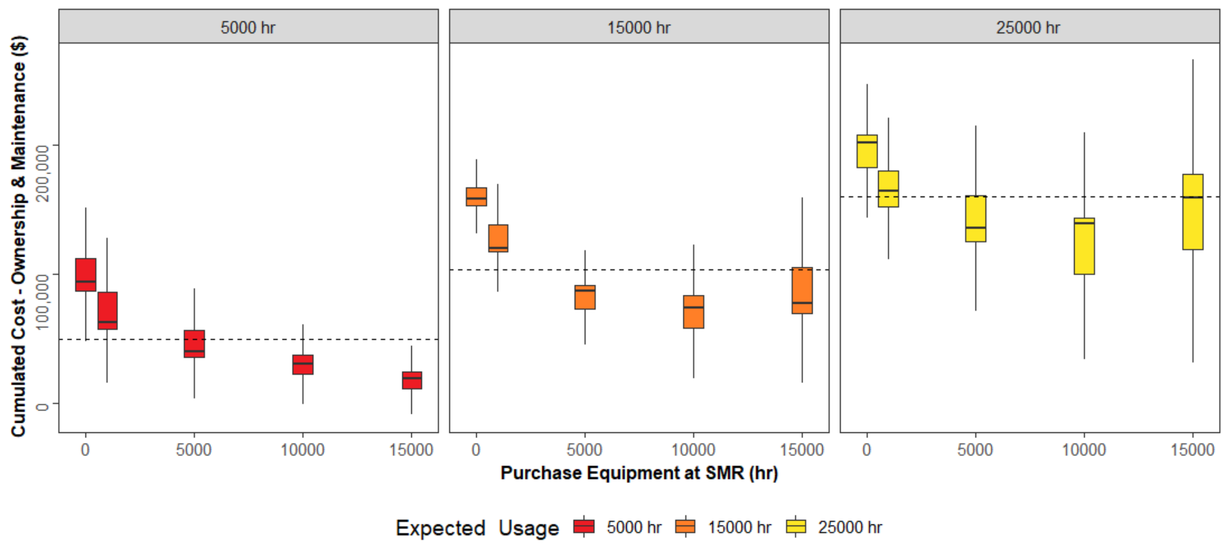


Figure 3.12 Cumulative Cost for Equipment Purchased at Various SMR

Furtheremore, Figure 3.13 illustates the cumulative cost per engine hour (average cost) for 3 expected usages given 5 scenarios, which demonstrate a similar pattern. Eventually, this case study (along with its associated discussions) are validated by construction equipment experts from the industry partner organization and the experience-based conclusions match with the analysis results.

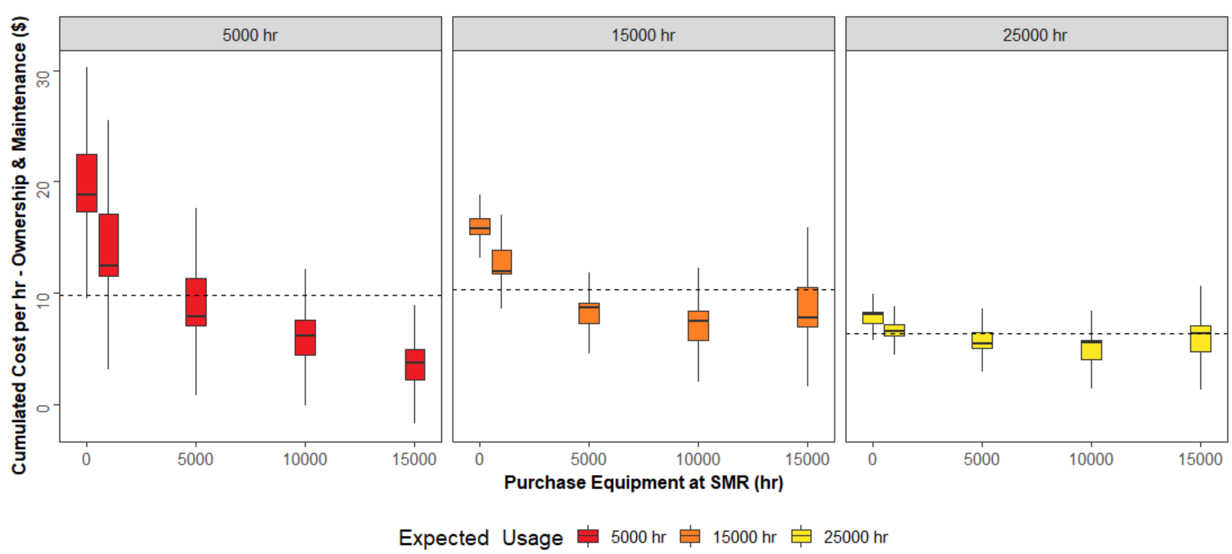


Figure 3.13 Cumulative Cost per Hour for Equipment Purchased at Various SMR

3.6 Potential Industry Applications

Lack of a robust cost-analysis model for equipment management can make the purchasing and selling decision-process challenging in practice. The proposed method is designed to quantify the range of cumulative cost at any engine hour (SMR) to provide decision support. In the case study, the proposed method is applied to compare the life-cycle cost of equipment purchased at different engine hours with various expected usage hours. Additionally, the proposed method can potentially be applied to address the following issues: (1) the break-even point (period) for equipment, and (2) equipment cash flow analysis.

Break-even Point (Period) for Equipment

Practitioners of the construction industry are interested in finding the revenue break-even point in a given piece of equipment's life-cycle in order to determine when to acquire or sell the equipment. With the consideration of uncertainty (made achievable through the proposed method in this paper) the break-even point becomes more meaningful to decision makers. Taking the Figure 3.14 as an example, a revenue line is obtained based on the fixed rate of return ($8.5/hr) and shown as the dashed line (Revenue=8.5·SMR). The median value of cumulative costs (combining ownership and maintenance costs) is represented by the black curve, while the upper limit and lower limit curves are represented in orange (calculations are conducted similar to Figure 3.10).

In this case, there are 3 break-even points (A, B, and C marked in red) in Figure 3.14. Accordingly, practitioners can utilize any of these break-even points in their decision making: point A (SMR_A)

can be used in an aggressive scenario, and the point C (SMR_C) for a conservative scenario. In practice, construction companies can rely on their historical data and the proposed method to develop their own graphs, similar to Figure 3.14, rather than using less reliable experience-based methods.

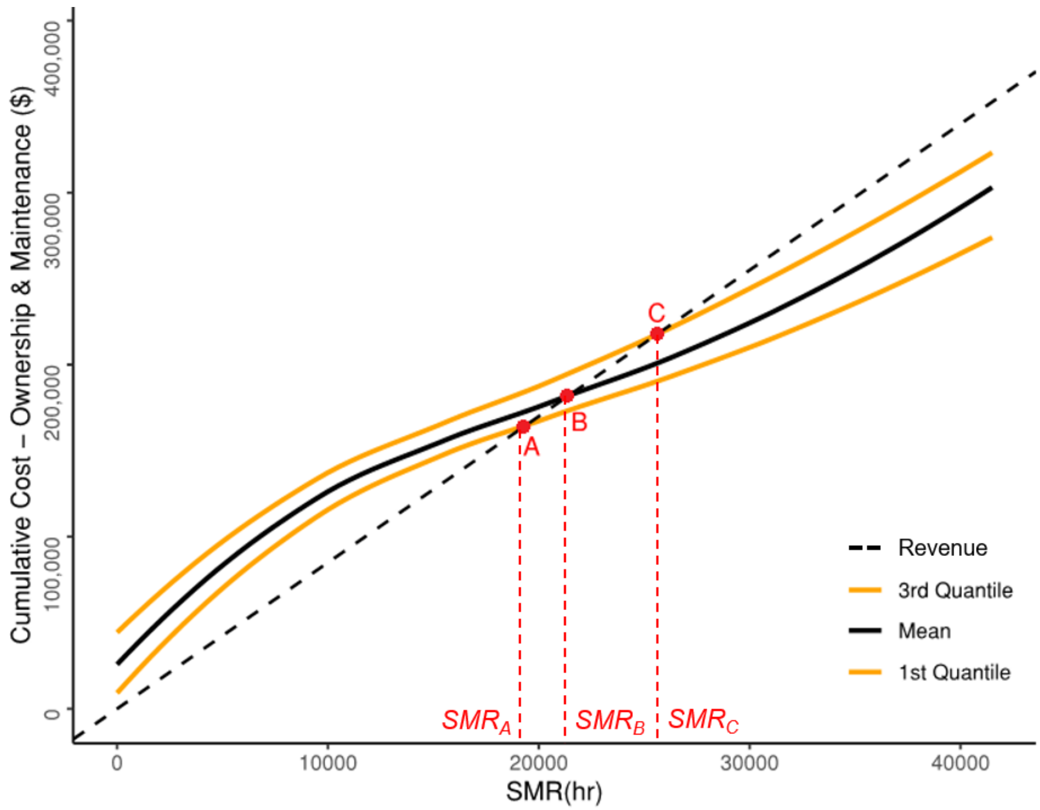


Figure 3.14 Breakeven Point for Equipment Cost

Equipment Cash Flow Analysis

Based on the proposed method, equipment costs can also be investigated through the cash flow perspective. In a hypothetical case a 32,000-hour lease agreement for equipment requires a premium paid at 15,000 hours as demonstrated in Figure 3.15. The cumulative cost curve is different: the cash flow before 15,000 hours (the shaded green area between 0 – 15,000 SMR) is positive. This scenario is probably more preferred by construction companies who do not have intents for holding equipment long-term, and instead seek short-term returns and cash flow. Similarly, construction practitioners can compare multiple finance methods using this approach to find the most meaningful scenario for equipment management.

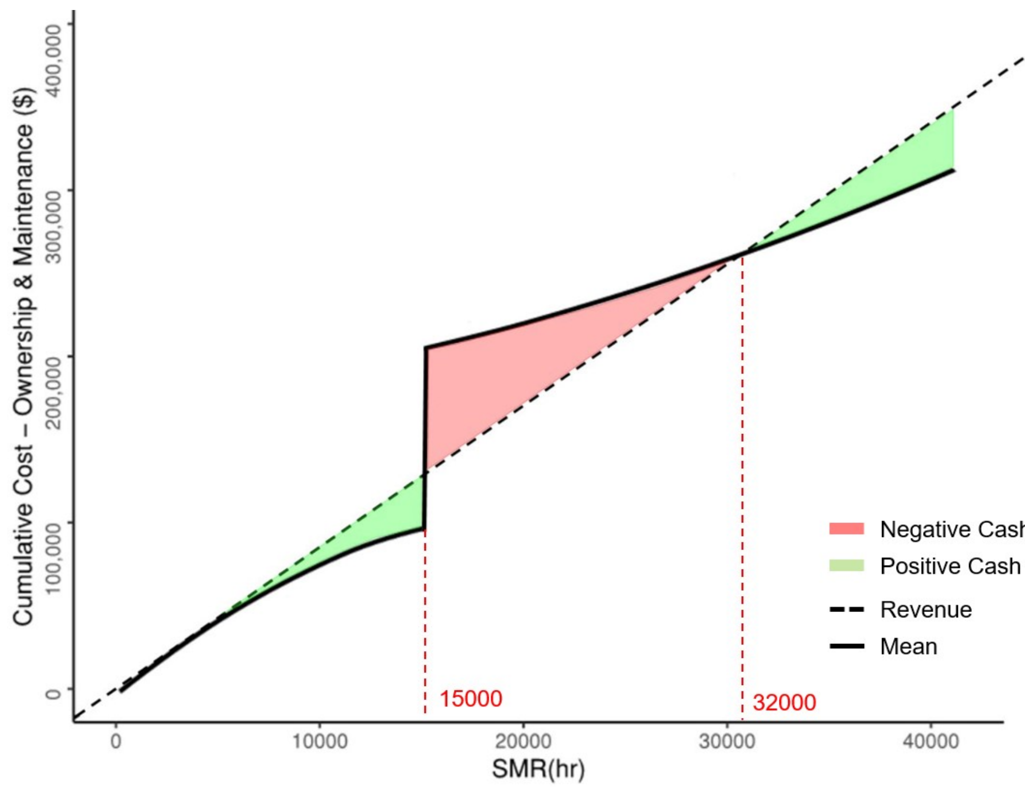


Figure 3.15 Illustration of Equipment Cost Analysis Based on Cash Flow

3.7 Conclusions and Future Work

This research proposes a data-driven, simulation-based analytics method for construction equipment management that incorporates K-means clustering and Expectation-Maximization (EM) methods to investigate the relationship and trends for ownership and maintenance costs. K-means clustering finds the turning points among maintenance stages for each equipment, and the EM algorithm quantifies the distribution of turning points based on the clustering results. The application of K-means clustering and the EM algorithm is provided in an illustrative example. The functionality of the data-driven simulation was demonstrated in a practical case-study. This research can be applied to satisfy the needs of large contractors in the aspects of both maintenance strategies and replacement strategies. Furthermore, the proposed method in this paper can be generalized to quantify uncertainties in similar problems.

To further support decision-making for equipment management and to facilitate the implementation of the proposed method into real practice, the following improvement of this research is required:

1) Maintenance cost and ownership cost are connected and accumulated based on the SMR in this study, which are normally collected when the maintenance or repair occurs. However, some construction companies may not keep track of the SMR (Mitchell 1998). To overcome the practical difficulties in date preparation, such as missing data, a scientific interpolation method needs to be developed.

2) Fixed costs (insurance, licensing, fuel, oil, and lubricant costs) do not account for large uncertainties, and therefore are excluded in the proposed method. It may be worthwhile to take

these factors into account so that the life-cycle cost of equipment can be more comprehensive and convincing.

3) Overall, equipment management is an economic analysis problem where the budget is constrained with internal and external impact factors. In the long term, the company is looking for a purchasing strategy; more importantly, they seek an economical strategy to incorporate cash flow and capital gains/losses. The data analytics research can further develop robust economic models to quantify the overall gain and loss.

3.8 Acknowledgements

This work was generously supported by Graham Industrial Services LP and was funded by a Collaborative Research and Development Grant (CRDPJ 492657) sponsored by the Natural Science and Engineering Research Council of Canada (NSERC). The authors would like to acknowledge Dr. Catherine Pretzlaw for her assistance in the technical writing

3.9 References

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

Amini, A.A., Mashayekhi, M., Ziari, H. and Nobakht, S. (2012). "Life cycle cost comparison of highways with perpetual and conventional pavements." *International Journal of Pavement Engineering*, 13(6), 553-568.

Bayzid, S.M., Mohamed, Y., and Al-Hussein, M. (2016). "Prediction of maintenance cost for road construction equipment: a case study." *Canadian Journal of Civil Engineering,* 43(5), 480-492.

Biller, B. and Gunes, C. (2010). "Introduction to simulation input modeling." *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, Maryland, USA

Chan, A., Keoleian, G. and Gabler, E. (2008). "Evaluation of life-cycle cost analysis practices used by the Michigan Department of Transportation." *Journal of Transportation Engineering*, 134(6), 236-245.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological),* 39(1), 1-22.

Drinkwater, R.W. and Hastings, N.A. (1967). "An economic replacement model." *Journal of the Operational Research Society*, 18(2), 121-138.

Er, K. C., Fernando, S., AbouRizk, S., and Ruwanpura, J. (2000). “Selecting the best tunnel construction for the 1st stage of the SESS tunnel using simulation.” *7th Annual Construction Research Forum*, Edmonton, Canada.

Fan, H., AbouRizk, S., Kim, H., and Zaïane, O. (2008). "Assessing residual value of heavy construction equipment using predictive data mining model." *Journal of Computing in Civil Engineering,* 22(3), 181-191.

Forgy, E.W. (1965). "Cluster analysis of multivariate data: efficiency versus interpretability of classifications." *Biometrics*, 21(1), 768-769.

Ghadam, P., Mehdi Ravanshadnia, and Saeed Ramezani. (2012). "Determining Economic Life of Earth Moving Equipment by Using Life Cycle Cost Analysis: Case Study." *Proceedings of International Conference on Construction and Real Estate Management*, 2012, 1-6.

Gransberg, D.D. and O′Connor, E.P. (2015). *Major equipment life-cycle cost analysis*. Minnesota Department of Transportation, Research Services & Library. Minnesota Department of Transportation, St. Paul, MN

Hajjar, D. and AbouRizk, S. (1997) "AP2-Earth: A simulation based system for the estimating and planning of earth moving operations." *Proceedings of the 1997 Winter Simulation Conference*, Atlanta, Georgia, USA.

Halpin, D.W. (2010). *Construction Management*. John Wiley & Sons, Hoboken, NJ.

Hastings, W.K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1), 97-109.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: Concepts and techniques.* Elsevier, Waltham, MA.

Ji, W., AbouRizk, S.M., Zaïane, O.R. and Li, Y. (2018). "Complexity analysis approach for prefabricated construction products using uncertain data clustering." *Journal of Construction Engineering and Management*, 144(8), 04018063.

Kuhl, M. E., N. M. Steiger, E. K. Lada, M. A. Wagner, and J. R. Wilson. (2009). "Introduction to modeling and generating probabilistic input processes for simulation." *Proceedings of the 2009 Winter Simulation Conference*, Austin, Texas, USA

Law, A. M. (2007). *Simulation Modeling and Analysis.* McGraw-Hill Education. New York City, NY.

Leemis, L. M. (2004). "Building credible input models." *Proceedings of the 2004 Winter Simulation Conference*, Piscataway, NJ.

Lucko, G. (2003). *A Statistical Analysis and Model of the Residual Value of Different Types of Heavy Construction Equipment*, Doctoral Dissertation. Virginia Tech, Blacksburg, VA.

Lucko, G., Anderson-Cook, C. M., and Vorster, M. C. (2006). "Statistical considerations for predicting residual value of heavy equipment." *Journal of Construction Engineering and Management*, 132(7), 723-732.

Mandapaka, V., Basheer, I., Sahasi, K., Ullidtz, P., Harvey, J.T. and Sivaneswaran, N. (2012). "Mechanistic-empirical and life-cycle cost analysis for optimizing flexible pavement maintenance and rehabilitation." *Journal of Transportation Engineering*, 138(5), 625-633.

Mitchell Jr, Z. W. (1998). *A Statistical Analysis of Construction Equipment Repair Costs using Field Data & the Cumulative Cost Model*, PhD Dissertation. Virginia Tech, Blacksburg, VA.

Mitchell, Z., Hildreth, J. and Vorster, M. (2011). "Using the cumulative cost model to forecast equipment repair costs: Two different methodologies." *Journal of Construction Engineering and Management*, 137(10), 817-822.

Mithraratne, N. and Vale, B. (2004). "Life cycle analysis model for New Zealand houses." *Building and Environment*, 39(4), 483-492.

Peña-Mora, F., Han, S., Lee, S., and Park, M. (2008). “Strategic-operational construction management: Hybrid system dynamics and discrete event approach.” *Journal of Construction Engineering and Management*, 134(9), 701–710.

Peurifoy, R., Ledbetter, W. and Schexnayder, C.J., 2002. *Construction, planning, equipment, and methods*, McGraw-Hill Education. New York City, NY.

Puri, V. and Martinez, J.C. (2012). "Modeling of simultaneously continuous and stochastic construction activities for simulation." *Journal of Construction Engineering and Management*, 139(8), 1037-1045.

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, available at: https://www.R-project.org/ (accessed April 5th, 2019).

Santos, J. and Ferreira, A. (2013). "Life-cycle cost analysis system for pavement management at project level." *International Journal of Pavement Engineering*, 14(1), 71-84.

Vorster, M. C. (2009). *Construction Equipment Economics.* Pen, Christiansburg, VA.

Vorster, M.C. (1980). *A systems approach to the management of civil engineering construction equipment*. PhD Dissertation. Stellenbosch University, Stellenbosch, South Africa.

Zong, Y. (2017). *Maintenance Cost and Residual Value Prediction of Heavy Construction Equipment*, MSc Thesis. University of Alberta, Edmonton, AB.

4 CHAPTER 4: EQUIPMENT LOGISTICIS PERFORMANCE MEASEUREMENT USING DATA-DRIVEN SOCIAL NETWORK ANALYSIS

4.1 Introduction

Equipment represents a large expense for heavy civil construction projects and corporations. While equipment management is essential for ensuring that projects are completed on time and on budget (Vorster 2009), optimization of this process can have a considerable impact on project efficiency and, in turn, overall cost. Accordingly, many construction organizations are interested in evaluating equipment management performance. In current practice, evaluation of equipment management is solely based on rates of equipment utilization, with increased utilization suggestive of improved management performance.

Equipment utilization represents slightly different concepts depending on the management level to which it is applied. Utilization at a project level (El-Rayes and Moselhi 2001; Wang et al. 2004) is based on equipment downtime, which is affected by a variety of factors such as site conditions, operator skills, equipment conditions, and force majeure (Prasad and Park 2004). Utilization at the corporate level, also known as deployment, assesses the utilization of equipment over its lifetime (Vorster 2009). This level of management, also known as centralized equipment management, is performed to allocate self-owned mobile equipment across various projects at a corporate level (Mitchell 1998; Fan et al. 2006). Factors influencing corporate-level equipment utilization include the length of the construction season, economic situation, and ongoing project numbers.

Although an important metric for evaluating equipment management performance, utilization rates do not consider the logistical effort associated with equipment management at a corporate level. Inefficient deployment of equipment between worksites and equipment shops can increase logistics-associated effort and expenditures and reduce the amount of time equipment is available to work. Methods capable of reliably quantifying logistical effort of equipment management practices, however, remain relatively unexplored. Since the social network analysis is a powerful tool examining interaction or relationship between studied objects, it can make effective use of existing company data of equipment logistics to investigate the equipment movements between projects and shops. Inspired by the social network analysis, the equipment movement between project sites and equipment shops can be investigated which has been rarely studied before.

Based on the social network analysis, this research has developed a novel decision metric—direct dispatch index (DDI)—that can be used to evaluate the logistical effort associated with equipment management practices. An innovative, social network analysis-based method for quantifying the DDI is proposed. Use of the DDI can enhance the ability of construction companies to more comprehensively and, in turn, more reliably assess equipment management practices and to compare them between various equipment managers or groups within the organization.

The content of this paper is organized as follows: First, previous research on equipment management and applications of social network analysis in construction are reviewed. Then, the DDI performance metric is formulated, data sources are detailed, data cleaning and fusion are described, and the social network theory and characteristics of social network analysis are introduced. To demonstrate functionalities of the decision-support metric, a case study is conducted using historical data collected by a large contractor in Edmonton, Canada. Potential applications of the DDI and strategies to improve performance are suggested. Research contributions, limitations, and future work are discussed.

4.2 Literature Review

Construction Equipment Management

Equipment management is an essential component of a construction business that can affect both project and corporate performance (Samee and Pongpeng 2016). Over the past decades, construction equipment management has been studied from various perspectives that can be classified into three major topics: (1) equipment costs associated with acquisition, operations, maintenance, and disposal of equipment (Mitchell 1998; Fan et al. 2008; Bayzid et al. 2016); (2) equipment and fleeting management by using the modern tracking technology (Azar and Kamat 2017); and (3) equipment selection and operation to improve equipment utilization at the project level (Chae and Yoshida 2010; Alshibani and Moselhi 2016).

Equipment cost is the most frequently studied amongst all equipment management factors. With the development of data mining technology, recent regression models capable of using historical equipment data to predict equipment maintenance cost at any point in time for any maintenance interval have been developed (Yip et al. 2014; Bayzid et al. 2016). Rather than using internal historical data, equipment residual value analysis is now primarily being performed using auction and re-sale records collected in online databases—a practice that is widely accepted in literature (Lucko 2003). Recently, advanced heuristic algorithms and spatial cost analysis based on regression models have been further developed (Lucko 2003; Fan et al. 2008; Ponnaluru et al. 2012). In short, these models are used to estimate life-cycle or maintenance costs of equipment, to provide analytical support for determining when equipment should be acquired, and to determine if a make or model is worth being acquired.

Due to the rapid development of tracking technology, recent equipment operations studies have focused on tracking and analyzing equipment data. Most of the GPS-based system developed by the equipment manufacturer such as the Caterpillar can locate individual heavy equipment and diagnose its mechanical health, such as equipment engine hours, fuel consumption, and geo-location of equipment, in real time, but these systems are challenged by lacking in-depth data analysis and accuracy (Azar and Kamat 2017). Taking advantage of pattern recognition technology and algorithm, recent research investigates more data such as the weight, payload and pose of the machine (Ibrahim and Moselhi 2014; Pradhananga and Teizer 2015).

At the project level, especially for material handling problem such as concrete delivery (Lu et al. 2007) and earthmoving projects (Song and Eldin 2012; Alshibani and Moselhi 2016), real-time tracking has also been used to record cycle time of equipment and actively reduce idle times on large worksites by facilitating fleet management and equipment dispatch. Taking advantage of shortest path algorithms in logistics, real-time optimization of transportation routes have been developed to improve earthmoving operations based on GPS on mining worksites (Choi and Nieto 2011). However, the overwhelming costs associated with achieving real-time data together with the rigidity of simulation and optimization methods have limited the practical application of these methods. Indeed, it remains common practice in construction to examine tracking data weekly or monthly, with analyses being conducted quarterly or annually.

While equipment logistics have been seldom studied or considered in equipment management research, several researchers have contended that consideration of equipment dispatch and transport may enhance decision-making (Fan et al. 2006; Hendi 2007). Furthermore, in contrast to real-time tracking data, equipment logistics data, which involves recording equipment movements from project to project, can be tracked economically and updated dynamically. In spite of these advantages, methods capable of transforming logistics data into useful information have not yet been reported.

Social Network Analysis

Social network analysis (SNA) was first introduced into sociology, anthropology, and political science based on knowledge from networks and graph theory. Nodes (i.e., vertices) and edges (i.e., ties) comprise the network structure of SNA, where nodes can represent individuals, groups, or companies and edges can represent relationships, communications, or movements between nodes.

Taking advantage of its visualization power, an increasing number of studies are adopting SNA to analytically evaluate social relationships and network characteristics (Zheng et al. 2016). SNA has been applied in construction engineering and project management, it has proven to be a powerful tool for illustrating business relationships and behaviors between construction projects (Borgatti and Foster 2003; Tortoriello et al. 2012; Hansen et al. 2005). Recently, human mobility of using civil infrastructures is studied by using geo-social network analysis through collect mobility data from Twitter. (Wang and Tyler 2015).

It is widely accepted that SNA can used to investigate network issues in the field of engineering project organization (Chinowsky and Taylor 2012). Among 63 recent SNA-based papers in construction engineering and management, 15 papers are categorized to be intra-organizational, 47 papers are categorized to be inter-organizational and only 1 paper is not categorized since the nodes are not individuals or organization but defects (Zheng et al. 2016). At the inter-organizational level, which comprises the scope of most studies, SNA is often used to evaluate business activities, such as supply chain management and strategic alliances. Many believe that the ability of SNA to conduct multiple-level analysis and integrate quantitative, qualitative, and graphical data represents a unique approach for solving certain project management problems (Ruan et al. 2012; Pryke 2012). SNA has been used to study (1) historical data for construction project coalitions, which revealed close relationships between some consultants and contractors

(Pryke 2004), and (2) resource management to investigate business relationships at the inter-organizational level (Sandhu and Helo 2006).

At the intra-organizational level, SNA can be applied to study the communication problems between key individuals in a complex network that are difficult to investigate using other methods. With the development of virtual design technologies in civil engineering, SNA has been conducted using digital logs generated by BIM software to visualize collaborations and rank importance of designers (Zhang and Ashuri 2018). In addition, intra-organizational SNA can demonstrate positive relations between relationships and performance, allowing the efficiency and performance of corporate operations to be evaluated using SNA (Lin and Tan 2013; Priven and Sacks 2015).

Given that resource management problem at the intra-organizational level has not yet been studied through SNA, this study will investigate the data obtained from one large general contractor, which may shed more insights into the practices of equipment management and find more opportunities to improve the performance. In brief, the past studies have not yet addressed the immediate needs of equipment management personnel by providing sufficient details to improve overall performance intra-organizationally. In order to overcome the identified limitations in previous research on the equipment management and social network analysis, the present research is intended to propose analytical method to improve the management which is ready for implementations.

4.3 Methodology

A data-driven performance measurement method capable of quantitatively and reliably assessing logistical effort associated with equipment deployment is proposed. Workflow of the methodology developed is summarized in Figure 4.1. Briefly, data collected from equipment tracking and project management systems are first fused and cleaned. Based on the mapped equipment logistics data, a social network model is established using social network theory. To map the management scope of each equipment shop, community structures are detected in the network using the Louvain method (Blondel et al. 2008). Centrality measurements are calculated to identify the importance of shops. The novel DDI performance metric is then used to evaluate the equipment logistics performance of each shop.

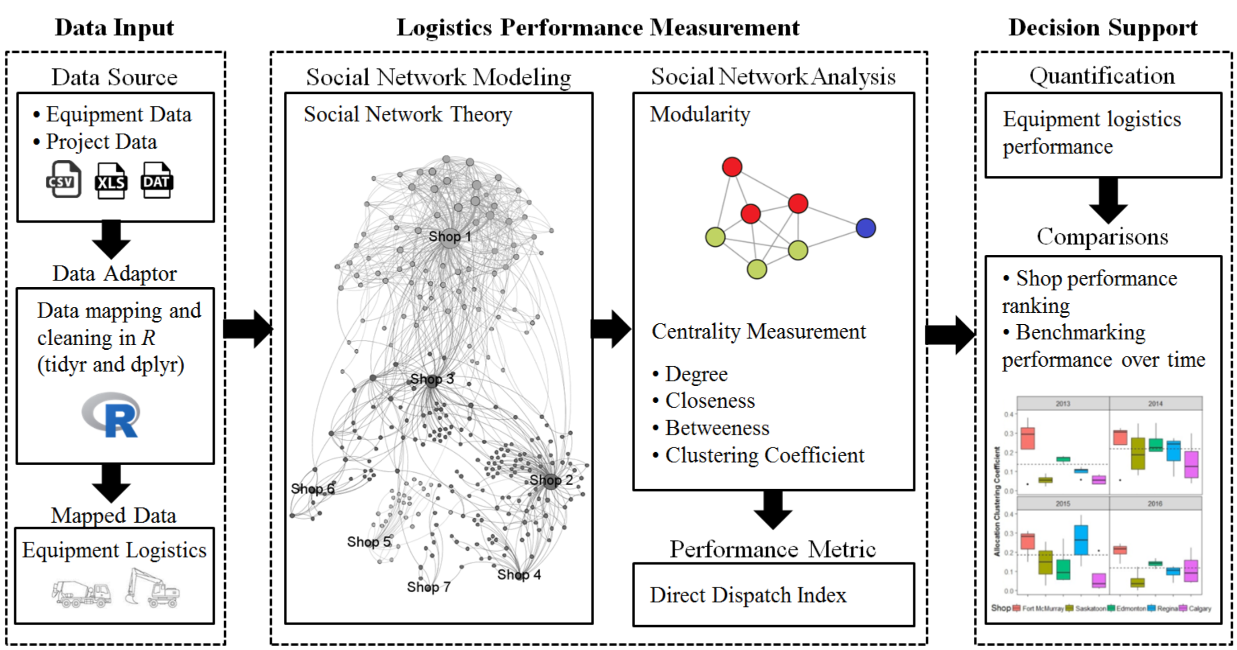


Figure 4.1 Methodology Flowchart of Data-Driven Measurements Method

Data Input

Equipment logistics, involving equipment dispatch and transport, can be extracted from equipment movement data, especially for heavy mobile equipment such as large excavators. Equipment manufacturers offer subscriptions that allow companies to record the real-time geo-locations of their equipment. Alternatively, self-developed systems, commercial tracking systems, or manual recording practices can be used to obtain real-time equipment location data. An example of equipment tracking data is shown in Table 4.1.

Table 4.1 Sample of Equipment Tracking Data

Equipment ID	Description	Location	Start Date	End Date	Utilization
10005217	Excavator	S600016	5/1/2013	5/31/2013	0.00
10005218	Excavator	PE12110	5/1/2013	5/31/2013	0.82
10005231	Excavator	PE11077	5/1/2013	5/31/2013	0.76

To connect the geo-locations with internal identifications of projects and shops, project details from the project management system are extracted. Details, such as project internal identifications, project location, and project descriptions, are used to geographically associate each piece of equipment with the equipment tracking data. Data adapters are then used to combine data from these various sources into one centralized dataset, providing richer information for data mining (Ji and AbouRizk 2018).

Data adapter integrate data from various data sources which could be relational databases into one centralized dataset (Ji and Abourizk 2018). In this study, the data adapter is designed to extract only the portion of the information contained in the equipment tracking and project management systems that are required for input into the social network model. The data adapter is developed in *R* (R Core Team 2019), an open-source statistics software program capable of handling large-scale data, which performs several functions including data connection, wrangling, cleaning, and mapping. In this study, the *dplyr* and *tidyr* packages in *R* were used to perform data wrangling tasks, which can process large-sized dataset in a relatively little time. Output of the data adapter are mapped logistics data, exemplified in Table 4.2, which detail the location and assigned project for each piece of equipment at any point in time.

Table 4.2 Sample of Mapped Equipment Logistics Data

Date	Equipment ID	Departure	Departure Location	Arrival	Arrival Location
5/13/2013	10005237	S600016	City A	PE13008	City B
5/15/2013	10013234	S600002	City C	S600016	City A
5/18/2013	10016007	PE11077	City D	PE11079	City E

Social Network Modeling

The social network model, presented as a weighted network diagram, consists of nodes and edges. In social network studies, nodes are, typically, the individual or group being studied, while edges represent relationships, communications, or behaviors between nodes. In this study, project and equipment shops, where equipment is used, maintained, and stored, are the individuals being studied. Similarly, the movement of equipment can be represented as a communication between a shop and a project, or between two project sites.

As such, nodes represent locations of project jobsites or equipment shops (i.e., storage sites), and edges represent the movement of equipment between sites. In the example outlined in Table 4.3, equipment shops (e.g. S600016 and S600019) and projects (e.g. PE09190, PE09194, and PE09196) are identified as nodes. The edge (e.g. ID 1), represents the movement of construction equipment between nodes S600016 and PE09190. Weights of edges are often used to denote the strength of the relationship between nodes. Here, the weight, w_{ij}, is defined as the total count of the equipment movements between node *i* and *j*, which can be obtained through accumulating each movement associated with individual equipment from historical logistics data. In this example, there are 7 movements along edge 1 between nodes S600016 and PE09190.

Table 4.3 Example of Logistics Data for Modeling the Social Network

Edge ID	Node *i*	Node *j*	Weight
1	S600016	PE09190	7
2	PE09196	PE09190	3
3	S600019	PE09190	1
4	S600016	PE09194	3
5	PE09190	PE09194	1
6	PE09196	PE09194	3

Social Network Analysis

A social network model can be used to (1) recognize patterns of relationships, (2) identify the importance of each node through centrality measurements, and (3) detect the community structure in a network as subgroups or subsets (Zhang and Ashuri 2018). Since shops and projects are usually divided into geographically-distinct groups in practice, the community structure can be detected by modularity to identify the management area of an equipment shop. The centrality of nodes representing the equipment shops can also be measured by either a social network model or a community structure of the social network.

Centrality measurements are used to rank the importance of nodes. Characteristics including degree (i.e., number of edges connected to the node in the network), closeness (i.e., average length of the shortest path between the node and all other nodes), betweenness (i.e., number of times that a node is on the shortest path between two other nodes in the network), and clustering coefficients (i.e., degree to which nodes in a social network tend to cluster together) are well-defined in SNA to quantitatively evaluate the centrality of a node. In an undirected graph, the local clustering coefficient can be calculated using Eq. 4.1, which ranges from 0 to 1.

$$C_i = \frac{2e_i}{k_i(k_i-1)} \tag{4.1}$$

where k_i is the number of neighbours of the ith node, and e_i is the number of connections between these neighbours.

Long-distance transport of equipment across states or provinces is a costly endeavor. Consequently, equipment management is usually divided geographically and assigned to equipment shops instead of managing the equipment at a corporate level. Each equipment shop manages the equipment in a certain region with management scope including planning, dispatch, maintenance, and repair. Accordingly, equipment movements within management regions often account for a substantial portion of the data, with a few equipment movements occurring from region to region to fulfill the urgent needs of projects. To investigate the performance of each shop, regions managed by each shop must be identified.

Modularity, which aims to divide networks into smaller groups and detect community structures within a network, can be used to identify the management region of each shop. Following the application of modularity techniques, the nodes in a community are more densely connected with each other that the rest of the network. Among the proposed modularity algorithms, the Louvain method, essentially a greedy optimization method, is the most popular due to its ability to outperform similar methods in terms of modularity and speed (Blondel et al. 2008). In each iteration of the Louvain method, the nodes are first grouped into small communities based on the value ΔQ as shown in Eq. 4.2.

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m}\right)^2\right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m}\right)^2 - \left(\frac{\sum_{tot}}{2m}\right)^2\right] \quad (4.2)$$

where $\sum_{in}$ is the sum of the weights of the links inside C, $\sum_{tot}$ is the sum of the weights of the links incident to nodes in C, k_i is the sum of the weights of the links incident to node i, $k_{i,in}$ is the sum of the weights of the links from i to nodes in C, and m is the sum of the weights of all of the links in the network.

Nodes in the communities then become the nodes, and the optimization method is again applied to the new network. Through iterations, modularity results are achieved, and tight relationships between projects and shops within the community structures of the network emerge. From this, the management scopes of shops are easily identified.

Direct Distance Index Performance Metric

Equipment dispatch in real practice must be investigated and divided into two strategies: (1) equipment that can be dispatched from the equipment shop or (2) equipment that can be dispatched from another project when the locations of two projects are close. Other than when equipment must be sent back to equipment shops for repairs or rebuilds, equipment should be dispatched from one project to another. Direct dispatch of equipment from project to project can not only reduce the logistical effort but may also reduce equipment travel time and, in turn, increase long-term equipment utilization.

Notably, clustering coefficients can be simply used to evaluate dispatch strategies by quantifying the number of times that equipment is dispatched from projects as a ratio of the number of dispatches from shops. For nodes representing shops, the numerator of the clustering coefficient (i.e., e_i) refers to the number of equipment movements from project to project. The limitation of the clustering coefficient is that the cost and time based on the distance has been ignored. While the number of movements is significant, it is also important to consider the distances associated with each movement. As such, to overcome the limitation of clustering coefficient, a second weight, distance, is introduced to achieve the DDI.

The DDI is, therefore, designed to quantify the reduction in the distance that equipment must travel to reach its destination. Here, the distances between nodes in kilometers, d_{ij}, are the distances that must be driven to transport equipment from one location to another given the geo-locations of the nodes for both shops and projects. In this study, driving distances are determined using navigation tools from transportation routes; algorithms to determine distances are not applied. *DDI*, which considers both the number of equipment movements, w_{ij}, and the logistical distances, d_{ij}, can be calculated using Eq. 4.3 and 4.4.

$$d_{ij}^{*} = d_{is} + d_{sj} \tag{4.3}$$

$$DDI = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(d_{ij}^{*} - d_{ij}) \times w_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^{*} \times w_{ij}} \tag{4.4}$$

where d_{ij} is the distance from node i to node j, d_{ij}^{*} is the distance from node i to node j through shop S, and w_{ij} is the total count of equipment movements between node i to node j.

The index ranges from 0 to 1, with a greater value indicating improved dispatch efficiency. When the index is 0, equipment is inefficiently dispatched from equipment shops to projects, and when the index is 1, all equipment is efficiently dispatched from project to project.

4.4 Illustrative Example

This example compares two dispatch plans, which are illustrated as social network models in Figures 4.2 and 4.3. Both social network models include a shop and four projects that are represented by five nodes. Shop and project locations, and, consequently, the distance between nodes (i.e., d_{ij}), are the same in both models. However, because equipment dispatch plans vary between the models, the number of equipment movements (i.e., w_{ij}) differs. Data are summarized in Tables 4.4 and 4.5.

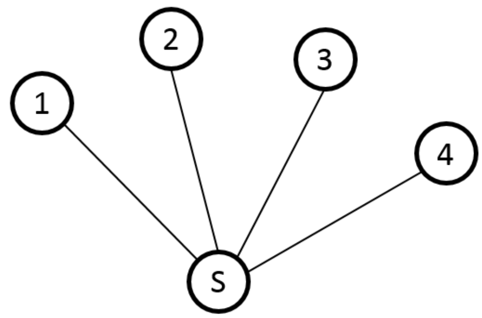


Figure 4.2 Social Network Model A

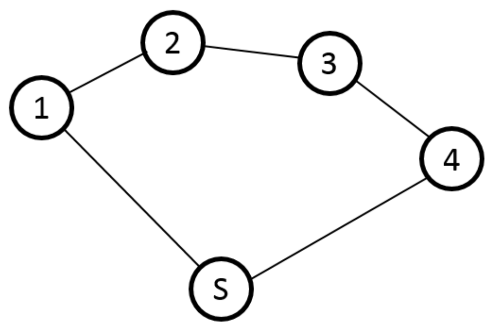


Figure 4.3 Social Network Model B

Table 4.4 Edge Data in Social Network Model A

Edge ID	Node *i*	Node *j*	w_{ij}	d_{ij}
1	S	1	10	10
2	S	2	10	10
3	S	3	10	10
4	S	4	10	10
5	1	2	0	1
6	2	3	0	1
7	3	4	0	1

Table 4.5 Edge Data in Social Network Model B

Edge ID	Node *i*	Node *j*	w_{ij}	d_{ij}
1	S	1	5	10
2	S	2	0	10
3	S	3	0	10
4	S	4	5	10
5	1	2	5	1
6	2	3	5	1
7	3	4	5	1

As per Eq. 4.2, clustering coefficients are calculated for Models A and B as 0 (=0/6) and 0.5 (=3/6), respectively. While there are a maximum of 6 edges between nodes 1, 2, 3, and 4, there are 0 edges in Model A and 3 edges in Model B. Using Eq. 4.3 and Eq. 4.4, *DDI* are calculated as 0 (=0/400) and 0.7125 (=19×5/400+19×5/400+19×5/400=285/400) for Models A and B, respectively, which are also explained in details in Appendix B. Note that denominators of both *DDI* ($\sum_{i=1}^{n} d_{ij}^{*} \times w_{ij}$) are the same (i.e., 400). For Model A, both the clustering coefficient and *DDI* of node *S* are 0. For Model B, the clustering coefficient of node *S* is 0.5, and the direct dispatch index is 0.7125. A greater *DDI* indicates that fewer detours comprise the dispatch plan outlined in Model B compared to that in Model A, consistent with illustrations in Figures 4.2 and 4.3, respectively.

As the number of projects managed by the shop increases, the difference between the clustering coefficient and DDI will increase. An extreme case, demonstrated in Figure 4, is illustrated. Here, the same dispatch plan in Model B is used to sequentially complete *N* projects in the same location that are managed by one shop. In this case, the clustering coefficient is 2/*N*, which approaches 0 when *N* is large. In contrast, however, the *DDI* remains close to 1. Altogether, these results demonstrate that dispatch efficiency can be reliably evaluated regardless of project number when using the DDI metric.

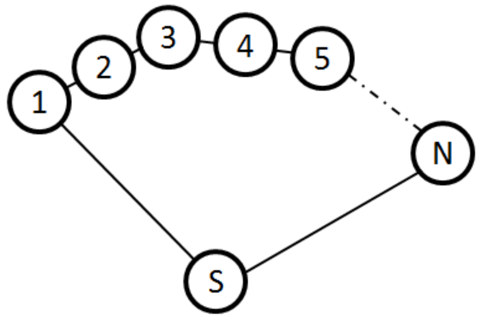


Figure 4.4 Extreme Example of Equipment Dispatch

4.5 Case Study

The proposed methodology was applied to a case study. Historical data were collected from a construction contractor in Alberta, Canada, between 2013 and 2016. Equipment data were extracted from the internal equipment management system, *SAP ERP* (SAP SE 2018), and project data were collected from a self-developed project management system. Data from both systems were combined, and missing data were omitted using *R* (R Core Team 2019).

Following the application of the proposed methodology, the social network model, based on the historical data, is demonstrated in Figure 4.5 and Figure 6. Note that all edges are bidirectional. Here, nodes denote project sites or equipment shops, the sizes of the nodes their degree, and the thicknesses of the edges are determined by the number of equipment movements (i.e., w_{ij}).

The social network model is comprised of 297 nodes including 7 equipment shops and 290 projects. The geo-location of each shop is illustrated in Figure 4.5. Based on the Louvain method, four communities (i.e., management areas) are detected and marked as orange (Shop 1), green (Shop 6/3), blue (Shop 5), and purple (Shop 7, 4, and 2) in Figure 4.6.

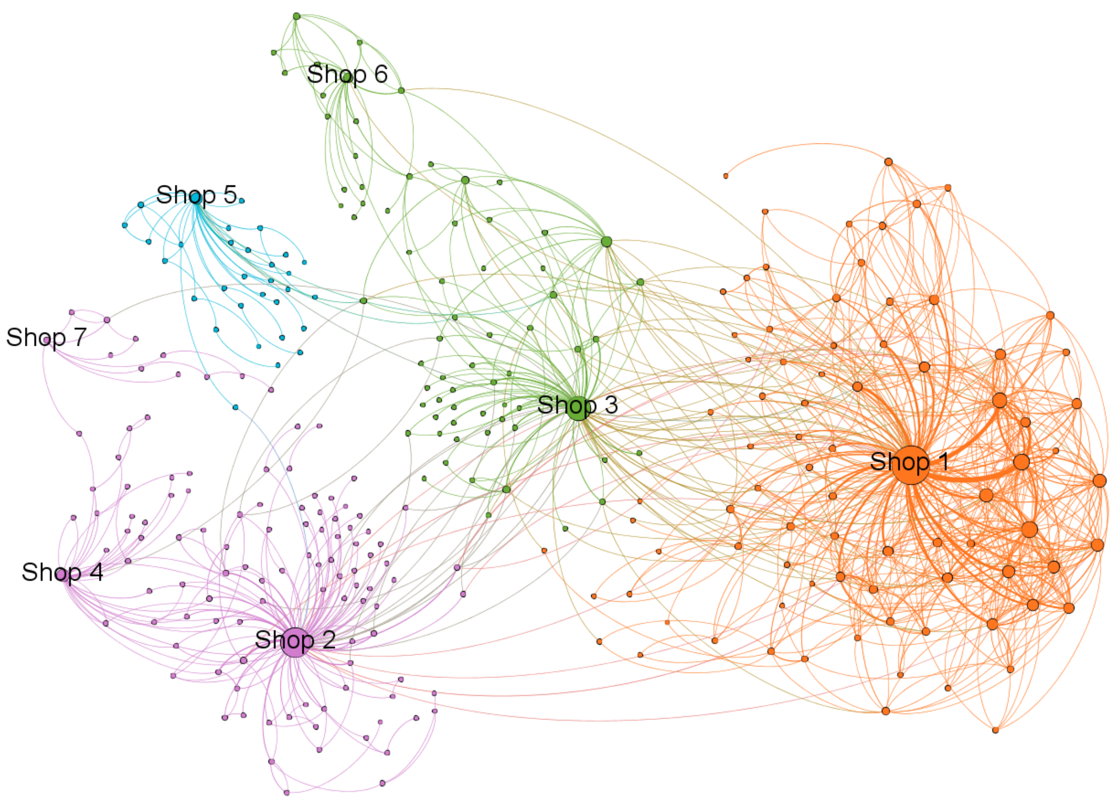


Figure 4.5 Social Network Model of Equipment Logistics after Modularity

In two communities, orange and blue, only one equipment shop was determined to be managing projects in that region. In the other two communities, green and purple, multiple shops were found to be responsible for the projects. One possible explanation is that small shops did not have sufficient equipment to supply projects, thereby requiring assistance from larger shops. In this case study, the resolution value of the Louvain method applied was 1; adjusting this value will affect the number of communities that are detected with the numbers of communities detected increasing as the resolution value increases.

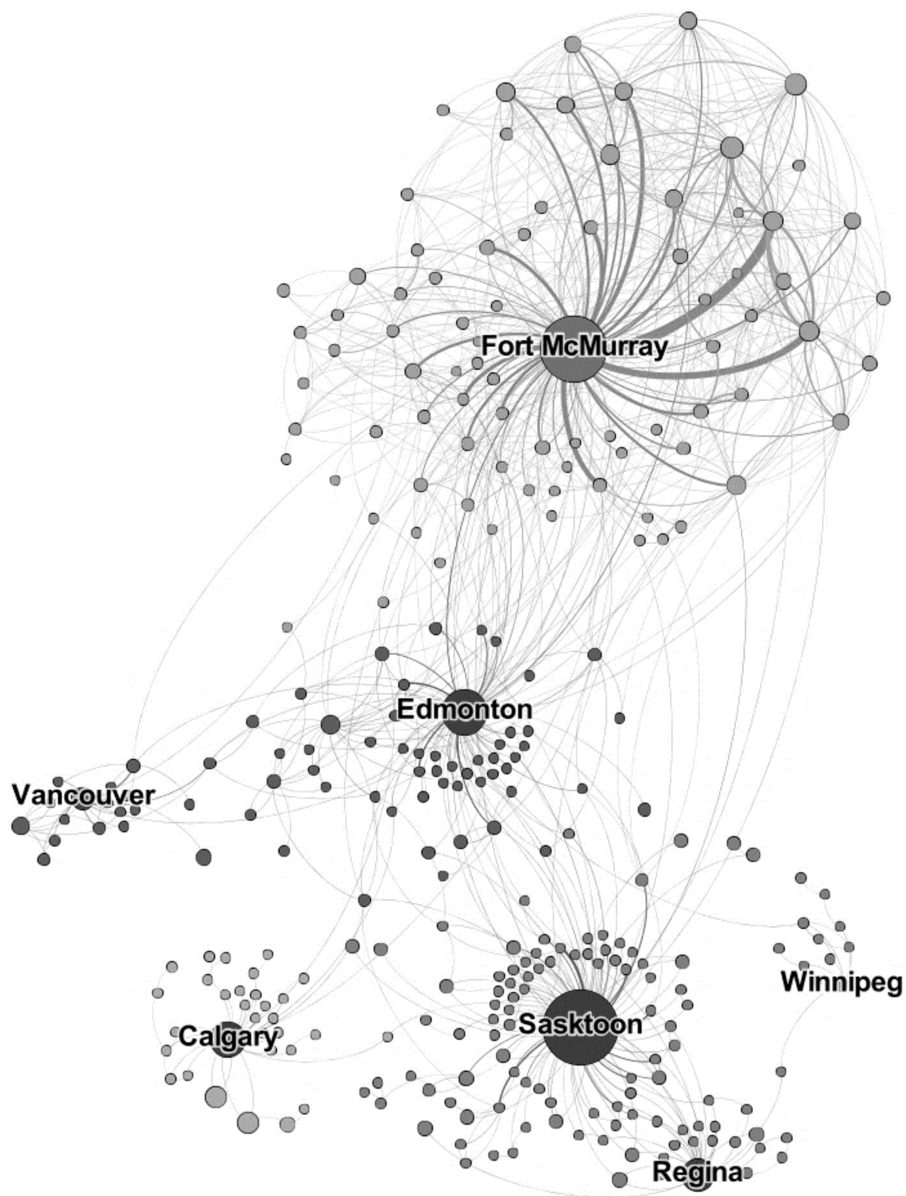


Figure 4.6 Social Network Model of Equipment Logistics

The five nodes representing five major shops are listed in Table 4.6 in order of degree (i.e., number of projects connected with the shop in the network). After dividing the network into communities, closeness, betweeness, the clustering coefficient, and DDI were calculated for each node (Table 4.6). The DDI was calculated based on four years' worth of data, which were used to evaluate the overall dispatch performance during this time.

Table 4.6 Logistical Data of Each Shop for Four Years

Node ID	Description	Degree	Closeness	Betweenness	Clustering Coefficient	DDI
S001	Shop 1	130	0.482	12193	0.118	0.254
S002	Shop 2	96	0.492	20678	0.016	0.066
S003	Shop 3	76	0.485	15222	0.038	0.155
S004	Shop 4	31	0.375	5305	0.035	0.107
S005	Shop 5	25	0.339	8152	0.012	0.038

The proposed methodology was then applied to each of the four years. Results of the five shops are illustrated in Figure 4.7, with the dashed line indicating the average value for each year.

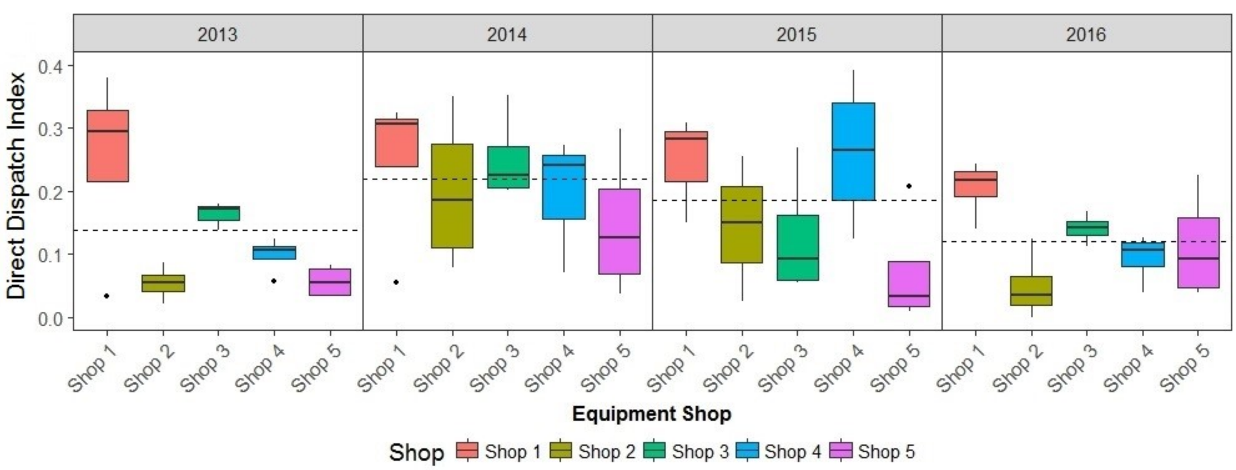


Figure 4.7 DDI for Each Shop by Year

Given the DDI, equipment management performance of the shops were evaluated and ranked, with a greater DDI indicative of a greater dispatch efficiency. Annual DDI of each shop were compared with the average four-year value. For example, the DDI for Shop 1 in 2016 was 0.212, which is above average for the five shops that year. However, its performance in 2016 was lower than that in 2015, which may require additional investigation. Performances of various equipment shops were also compared with each other.

Equipment Utilization Rate

As mentioned previously, equipment utilization is the primary metric by which equipment management performance is evaluated. It is generally accepted that high equipment utilization rates indicate efficient use of equipment. At a corporate level, equipment utilization, also known as deployment (Vorster 2009), can be defined by Eq. 4.5.

$$Equipment\ Utilization = \frac{Used\ Time}{Total\ Ownership\ Time} \tag{4.5}$$

where *Total Ownership Time* is the time (in days or hours) that equipment has been owned by the corporation and *Used Time* is the time (in days or hours) that equipment is allocated to a project regardless of its operation on the jobsite.

For comparative purposes, annual equipment utilizations for the five major shops are illustrated in Figure 4.8, with the dashed line indicating the average value for each year.

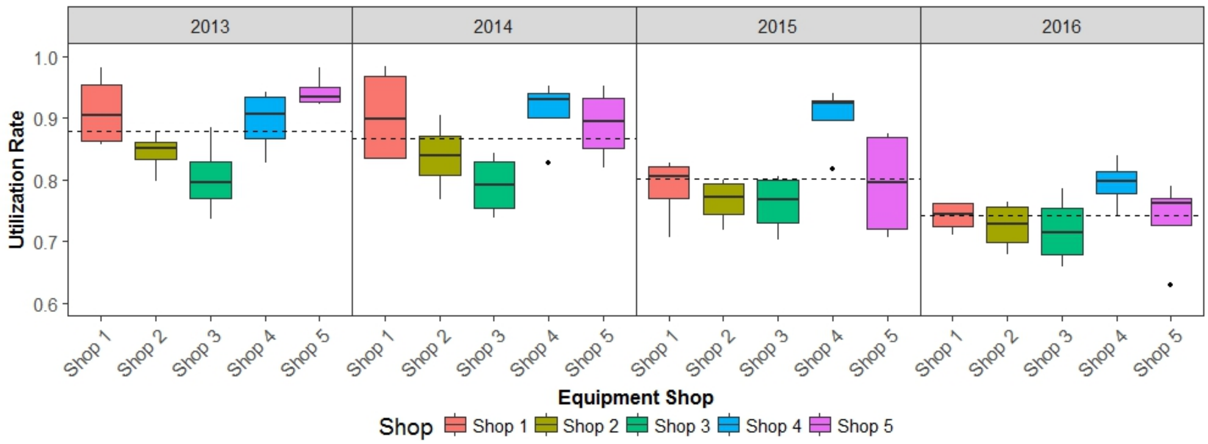


Figure 4.8 Equipment Utilization Rates of Each Shop by Year

Investigation of both metrics simultaneously can reveal further insight into the equipment management practices of each shop. While Shop 1 would have been considered average in terms of equipment utilization in 2016, its logistical performance was above average. Conversely, while Shop 2 is close to average with respect to its utilization rate in 2016, it is considerably below average with respect to logistical performance. Consideration of both factors reveals that, although these two shops share a similar utilization rate, Shop 1's equipment management performance exceeds that of Shop 2.

To ensure the proposed performance metric is capable of reliably evaluating the practical logistics performance, expert validation was conducted. Three equipment managers, each with more than ten-years working experience, were invited to evaluate the logistics performance of each shop based on their own professional experience and knowledge. They validated that results achieved based on the proposed DDI in the case study is aligned with the practical performance of equipment logistics management. The validation outcome demonstrates that feasibility and applicability of the proposed DDI.

4.6 Potential Applications

Lack of reliable, comprehensive performance measurements renders the improvement of equipment management challenging in practice. The proposed DDI performance metric is designed to quantitatively evaluate the logistical performance of equipment shops and to provide analytical decision-support to equipment managers and executives of construction companies. Potential applications of the proposed performance metrics include but not limited to the following areas: (1) benchmarking equipment management, (2) logistics-oriented equipment management, and (3) resource management.

Benchmarking Equipment Management

Following the implementation of the proposed methodology, executives will be able to determine which equipment shops are most proficient in equipment logistics. Managers of these equipment shops should be invited to share their professional knowledge for employee training purposes, particularly in the area of equipment dispatch. Companies may also standardize the equipment management process as per the high-performance equipment shops' dispatching strategies. Sharing the best practices company-wide are essential for improving overall equipment management.

Logistics-oriented Equipment Management

The DDI emphasizes the importance of considering equipment logistics in equipment management. Currently, logistical costs are not deliberately considered in equipment dispatch and transport. Examination of current practices using the proposed analytical method (or the development of an optimization method) can be used to design equipment dispatch plans that minimize logistical costs while maximizing utilization rates. Logistics-oriented equipment management offers companies the potential to improve long-term equipment management efficiency.

Resource Management

As a major resource in construction, it is anticipated that the proposed methodology can be generalized to other resource-based logistical problems. Social network theory and analysis can be easily generalized and embedded into the current material or labour management systems to visualize dynamic data and facilitate performance evaluation. Poor resource logistics can be identified and mitigated in a timely manner.

4.7 Conclusions and Future Work

Previous research has not yet addressed how to make use of equipment logistics data collected from equipment and project management systems to enhance decision-making. This research proposes the use of a social network analysis-based approach commonly applied in sociology to facilitate the visualization of equipment logistics and investigation of logistical performance. A dispatch distance-based performance metric, which can be used in conjunction with other metrics, such as the equipment utilization, to evaluate the performance of intra-organizational equipment management, is also proposed. The ability of the DDI metric to quantify the distance savings of various dispatch plans was demonstrated in the illustrated example provided, and the functionality of the proposed approach was confirmed following its application to a practical case study. When examined in conjunction with equipment utilization rates, the social network analysis-based method together with the DDI index can be used to more comprehensively examine equipment management practices.

Efficiency of equipment dispatch plans varies considerably between shops and organizations. It is possible to improve equipment management through benchmarking of the new performance metric. In addition, best practices identified using the proposed method can be shared within the company or even between companies to improve the time and cost-effectiveness of equipment dispatch. Research deliverables are anticipated to be of immediate use in practice and to satisfy the needs of large contractors that are eager to more comprehensively evaluate equipment management performance. Furthermore, the social network analysis approach described here can be generalized to identify and solve problems of other logistics-associated resources, such as labor and material.

To further support decision-making for equipment management and to facilitate the implementation of the quantitative, integrative methodology into real practice, further improvements of this research will be required:

1) The relationship between equipment utilization and the DDI may be further studied. Equipment utilization is primarily affected by the length of the construction season, ongoing project numbers, and economic conditions, which may not impact the DDI. In short, after collecting sufficient time series of both performance metrics, a serial correlation analysis can, and should, be conducted.

2) In this study, it is assumed that logistical costs are primarily determined by distance. However, the logistical costs may also be affected by the size of equipment, the remoteness of project locations, and the permit fee of certain routes. The replacement of the second weight in the DDI with these logistical costs may serve as a method for rapidly estimating cost savings associated with various equipment dispatch plans in contrast to the very time-consuming process of obtaining multiple quotations from equipment transportation companies.

3) Other than optimizing the equipment logistics, equipment dispatch can also be improved by adjusting the management region of equipment shops and the quantity of self-owned equipment. It is believed that the proposed performance metric can still quantitatively evaluate the dispatch efficiency after these improvements. However, the impact of improvement strategies on the performance metric should be studied and associated sensitivity analysis should be conducted.

4.8 Acknowledgements

This research is funded by the Natural Science and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Grant (CRDPJ 492657). The authors would like to acknowledge Jeffrey Gossain and Roman Gundyak for sharing their knowledge and expertise of equipment management, Dr. Yihan Zhao for sharing her knowledge in data visualization, and Graham Industrial Services LP for providing data.

4.9 References

Alshibani, A., and Moselhi, O. (2016). "Productivity based method for forecasting cost & time of earthmoving operations using sampling GPS data." *Journal of Information Technology in Construction (ITcon),* 21(3), 39-56.

Azar, E.R. and Kamat, V.R. (2017). "Earthmoving equipment automation: a review of technical advances and future outlook." *Journal of Information Technology in Construction (ITcon)*, 22(13), 247-265.

Bayzid, S. M., Mohamed, Y., and Al-Hussein, M. (2016). "Prediction of maintenance cost for road construction equipment: a case study." *Canadian Journal of Civil Engineering,* 43(5), 480-492.

Blondel, V. D., Guillaume, J., Lambiotte, R., and Lefebvre, E. (2008). "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment,* 2008(10), P10008.

Borgatti, S. P., and Foster, P. C. (2003). "The network paradigm in organizational research: A review and typology." *Journal of Management,* 29(6), 991-1013.

Chae, S., and Yoshida, T. (2010). "Application of RFID technology to prevention of collision accident with heavy equipment." *Automation in Construction,* 19(3), 368-374.

Chinowsky, P. and Taylor, J.E., 2012. "Networks in engineering: an emerging approach to project organization studies." *Engineering Project Organization Journal*, 2(1-2), 15-26.

Choi, Y., and Nieto, A. (2011). "Optimal haulage routing of off-road dump trucks in construction and mining sites using Google Earth and a modified least-cost path algorithm." *Automation in Construction,* 20(7), 982-997.

El-Rayes, K., and Moselhi, O. (2001). "Optimizing resource utilization for repetitive construction projects." *Journal of Construction Engineering and Management,* 127(1), 18-27.

Fan, H., AbouRizk, S., Kim, H., and Zaïane, O. (2008). "Assessing residual value of heavy construction equipment using predictive data mining model." *Journal of Computing in Civil Engineering,* 22(3), 181-191.

Fan, H., Kim, H., and Zaïane, O. R. (2006). "Data warehousing for construction equipment management." *Canadian Journal of Civil Engineering,* 33(12), 1480-1489.

Hansen, M. T., Mors, M. L., and Løvås, B. (2005). "Knowledge sharing in organizations: Multiple networks, multiple phases." *Academy of Management Journal,* 48(5), 776-793.

Hendi, A. (2007). "Decision support system for equipment selection in construction projects." *Masters Abstracts International*, 46(3).

Ibrahim M. and Moselhi O. (2014). "Automated productivity assessment of earthmoving operations". *Journal of Information Technology in Construction (ITcon)*, 19, 169-184.

Ji, W., and AbouRizk, S. M. (2018). "Simulation-Based Analytics for Quality Control Decision Support: Pipe Welding Case Study." *Journal of Computing in Civil Engineering*, 32(3), 05018002.

Lin, C., and Tan, H. (2013). "Performance measurement in the public sector: Example of the building administration authorities in Taiwan." *Journal of Management in Engineering,* 30(1), 97-107.

Lu, M., Chen, W., Shen, X., Lam, H.C. and Liu, J. (2007). "Positioning and tracking construction vehicles in highly dense urban areas and building construction sites. " *Automation in Construction*, 16(5), 647-656.

Lucko, G. (2003). *A Statistical Analysis and Model of the Residual Value of Different Types of Heavy Construction Equipment*, Doctoral Dissertation. Virginia Tech, Blacksburg, VA.

Mitchell Jr, Z. W. (1998). *A Statistical Analysis of Construction Equipment Repair Costs using Field Data & the Cumulative Cost Model*, Doctoral Dissertation. Virginia Tech, Blacksburg, VA.

Newman, M., Barabasi, A., and Watts, D. J. (2011). *The structure and dynamics of networks.* Princeton University Press, Princeton, NJ.

Ponnaluru, S. S., Marsh, T. L., and Brady, M. (2012). "Spatial price analysis of used construction equipment: the case of excavators." *Construction Management and Economics,* 30(11), 981-994.

Pradhananga N. and Teizer J. (2015). "Cell-based construction site simulation model for earthmoving operations using real-time equipment location data. " *Visualization in Engineering*, 3(12), 1-16.

Prasad Nepal, M., and Park, M. (2004). "Downtime model development for construction equipment management." *Engineering, Construction and Architectural Management,* 11(3), 199-210.

Priven, V., and Sacks, R. (2015). "Effects of the last planner system on social networks among construction trade crews." *Journal of Construction Engineering and Management,* 141(6), 04015006.

Pryke, S. (2012). *Social network analysis in construction.* John Wiley & Sons, Hoboken, NJ.

Pryke, S. D. (2004). "Analysing construction project coalitions: exploring the application of social network analysis." *Construction Management and Economics,* 22(8), 787-797.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, available at: https://www.R-project.org/ (accessed

May 18, 2018). X., Ochieng, E. G., Price, A. D., and Egbu, C. O. (2012). "Knowledge integration process in construction projects: a social network analysis approach to compare competitive and collaborative working." *Construction Management and Economics,* 30(1), 5-19.

Samee, K., and Pongpeng, J. (2016). "Structural equation model for construction equipment management affecting project and corporate performance." *KSCE Journal of Civil Engineering,* 20(5), 1642-1656.

Sandhu, M., and Helo, P. (2006). "A network approach to project business analysis." *Engineering, Construction and Architectural Management,* 13(6), 600-615.

SAP SE (2018). *Proven, time-tested on-premise ERP.* Toronto, Ontario, Canada, available at: https://www.sap.com/canada/products/enterprise-management-erp.html (accessed May 18, 2018).

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

Tortoriello, M., Reagans, R., and McEvily, B. (2012). "Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units." *Organization Science,* 23(4), 1024-1039.

Vorster, M. C. (2009). *Construction equipment economics.* Pen, Christiansburg, VA.

Wang, H., Zhang, J., Chau, K., and Anson, M. (2004). "4D dynamic management for construction planning and resource utilization." *Automation in Construction,* 13(5), 575-589.

Wang, Q. and Taylor, J.E., 2015. "Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms." *Journal of Computing in Civil Engineering*, 30(2), 04015004.

Yip, H., Fan, H., and Chiang, Y. (2014). "Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models." *Automation in Construction*, 38 30-38.

Zhang, H., and Li, H. (2004). "Simulation-based optimization for dynamic resource allocation." *Automation in Construction,* 13(3), 409-420.

Zhang, L., and Ashuri, B. (2018). "BIM log mining: Discovering social networks." *Automation in Construction,* 91 31-43.

Zheng, X., Le, Y., Chan, A. P., Hu, Y., and Li, Y. (2016). "Review of the application of social network analysis (SNA) in construction project management research." *Internal Journal of Project Management,* 34(7), 1214-1225.

5 CHAPTER 5: CONCLUSION

5.1 Research Conclusions

This research outlines the development of data-driven, simulation-based analytics to provide support for construction equipment management. It focuses on the costs of equipment, specifically maintenance and ownership costs, both separately and together, to fill the research gaps in earlier studies. In addition, the management performance of equipment logistics was studied through visualization and quantification for the first time, providing executive decision-makers supplementary analytical criteria other than equipment utilization to examine overall equipment management.

Chapter 2 proposes a dynamic, mathematical approach providing for decision-makers in equipment acquisitions and disposal that is capable of dynamically and continuously adapting to new market data. Two major parameters, namely bargain percentage and transaction quantity, are first proposed, and Bayesian inference is used to dynamically update parameters. The functionality and validity of the proposed method was demonstrated in a practical case study that examined transaction data for a popular hydraulic excavator, the CAT 320. Bargain percentage and transaction quantities were calculated from a “historical” portion of the data and were updated using “new” market data. Results (i.e., bargain likelihoods) obtained using the proposed method were found to strongly correlate with the actual number of bargains observed, indicating that the method was capable of generating valid, representative results.

Chapter 3 proposes a data-driven, simulation-based analytics method for construction equipment management to quantify the life-cycle cost of equipment and its uncertainty. The proposed method incorporates K-means clustering, for finding the turning points among maintenance stages for each type of equipment, and the Expectation-Maximization (EM) algorithm, for quantifying the distribution of turning points based on the clustering results. The application of K-means clustering and the EM algorithm is applied to an illustrative example and practical case-study. It can quantify uncertainties of equipment costs and satisfy the needs of large contractors for both maintenance and replacement strategies.

Chapter 4 proposes a social network analysis-based approach to facilitate the visualization of equipment logistics and investigation of logistical performance. A dispatch distance-based performance metric (DDI), which can be used in conjunction with other metrics such as equipment utilization to evaluate the performance of intra-organizational equipment management, is also proposed. The ability of the DDI metric to quantify the distance savings of various dispatch plans was demonstrated in the illustrative example provided, and the functionality of the proposed approach was confirmed following its application to a practical case study. When examined in conjunction with equipment utilization rates, the social network analysis-based method together with the DDI can be used to more comprehensively examine equipment management practices. Furthermore, the social network analysis approach described here can be generalized to identify and solve problems of other logistics-associated resources, such as labor and material.

5.2 Academic Contributions

This research contributes to new knowledge and innovation that adds value to the state of the art and its outcomes have resulted in several academic contributions：

- Applying the social network analysis to visualize equipment and project tracking data.
- Conceptualizing the equipment life-cycle cost problem into a data-driven, simulation-based analytical problem.
- Conceptualizing the equipment logistics problem into a social network model.
- Defining a novel performance metric for assessing equipment logistics performance.
- Integrating maintenance costs of heavy equipment with ownership costs, which have typically been studied separately.
- Examining quantity cost uncertainties in the equipment life-cycle cost, which can be further generalized into other similar life-cycle cost analysis problems.
- Introducing Bayesian inference for updating the historical data of equipment transactions with the newly-generated market data to realign with the dynamic, real-time data.
- Introducing a novel concept, likelihood of bargain, to quantify based on the relative value of equipment rather than on absolute RMV.
- Proposing an input modeling method, embedded within the K-means clustering and Expectation-Maximization methods, which can be generalized for other simulation problems.

5.3 Industrial Contributions

The industrial contributions from the collaborative research efforts are as follows:

- Alleviating the need for time-intensive redevelopments by end users. With continuously and dynamically updated data, practitioners can determine the optimal time, price, and location to purchase used equipment despite the tremendous amount of quantitative transaction data available.
- Facilitating the application of dynamic data-driven methods by practitioners who are not required to understand complicated models. The underlying mathematical complexity has been reduced for practitioners to understand, and in turn, to feel comfortable relying on the results of the model.
- Improving equipment management by adding performance metrics regarding equipment logistics and dispatch. When examined with equipment utilization rates, the social network analysis-based method together with the DDI index can be used to more comprehensively examine equipment management practices.
- Providing decision makers analytical support by accurately predicting which geographic locations were associated with the greatest likelihood of bargains.
- Satisfying the needs of large contractors to understand and quantify uncertainties in the life-cycle cost of construction equipment for both maintenance and replacement strategies.

5.4 Research Limitations

Although the research findings in above chapters support the developed approaches, limitations in this research should be noted. To facilitate the implementation of the quantitative, integrative methodology into real practice, further improvements of this research can be applied:

- Fixed costs – or operating costs (i.e. insurance, licensing, fuel, oil, and lubricants) – do not typically contain large uncertainties; thus, they have been excluded in the proposed method. It may be worthwhile to take these factors into account so that the life-cycle cost of equipment is more comprehensive and convincing.
- Logistical costs are assumed to be determined by distance. However, logistical costs may also be affected by the size of equipment, remoteness of project locations, and permit fees for certain routes. Replacing the second weight in the DDI with these logistical costs may serve as a method for rapidly estimating cost savings associated with various equipment dispatch plans, in contrast to the very time-consuming process of obtaining multiple quotations from equipment transportation companies.
- The equipment is assumed to be purchased instead of leased or rented. The ownership cost of leased or rented equipment is different. Its cash flow model is critical and should be studied in the further research.
- The data sources of maintenance costs should be further investigated and discussed. Acquiring maintenance cost data is time-consuming and hard to process and understand,

while ownership cost data is readily available. Some construction companies may not even keep track of the SMR, which increases the difficulty for acquiring the data (Mitchell 1998). To overcome the practical difficulties in data preparation, such as missing data, a scientific interpolation method should be developed.

- The relationship between equipment utilization and the DDI could be studied further. Equipment utilization is primarily affected by the length of the construction season, ongoing project numbers, and economic conditions, which may not impact the DDI. In short, after collecting sufficient time series of both performance metrics, a serial correlation analysis can, and should, be conducted.
- The trade-off between the time required for and the cost associated with data collection should be evaluated. Acquisition of dynamic data costs subscription fees or labor costs for manual data collection. The data sources should be further investigated by practitioners.

5.5 Future Directions

This section reveals possible future directions based on this doctoral research work.

- To develop a data-driven approach to obtain the break-even point (period) for equipment. Practitioners in the construction industry are interested in finding the revenue break-even point in a given piece of equipment's life-cycle to determine when to acquire or sell the equipment.

- To develop a data-driven, simulation-based approach to investigate equipment cash flow over time. When uncertainty is considered (through the proposed method in this thesis), whether the cash flow is positive or negative becomes more meaningful to decision makers.
- To generalize the proposed approach for equipment logistics to the other resource management problems. Social network theory and analysis could be generalized and embedded into current material or labour management systems to visualize dynamic data and facilitate performance evaluation. Poor resource logistics can be identified and mitigated in a timely manner.

Two more research topics could be studied in the future when more data from other systems (estimating systems, resource management systems, etc.) become available.

Data-Driven Equipment Usage Prediction

To properly manage equipment, understanding the equipment costs alone is not sufficient. Decision-makers need to understand future equipment demand and costs. Equipment usage hours over time is useful to equipment managers or decision-makers. However, the current empirical method is not reliable or analytical and could be improved.

Long-term predictions would provide a buffer for making decisions and a better idea of equipment usage fluctuations. Thus, forecasting equipment usage over the long-term and determining proper management strategies would be beneficial. Prediction of equipment usage, both short-term and long-term, has not been studied. The seasonal pattern of equipment hourly usage, illustrated in the Figure 5.1, was determined from a simulation-based prediction, but its accuracy has not been tested (Liu et al. 2018).

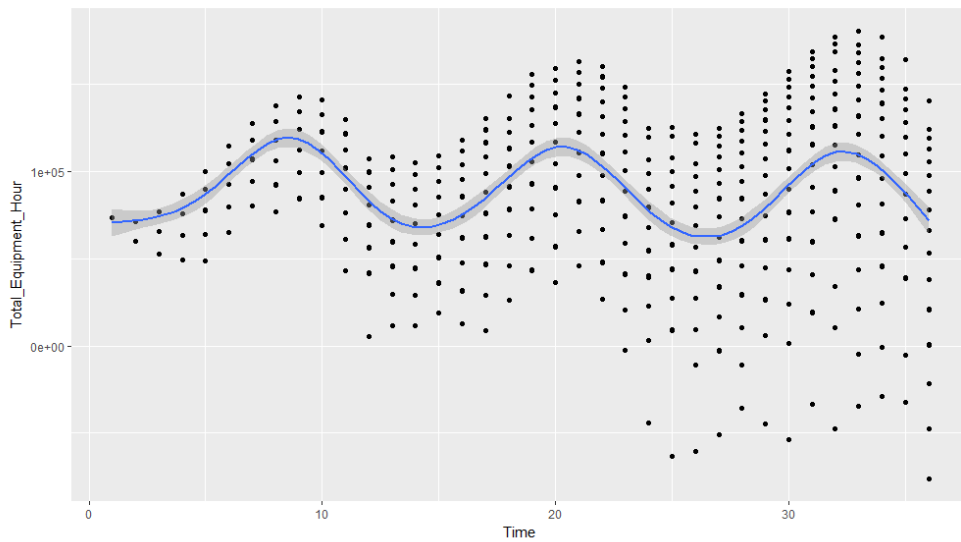


Figure 5.1 Long-term Prediction of Equipment Usage Hours

The estimating system has been used for pre-qualifications, bidding and change-orders of future projects. The data adaptor, which incorporates both estimating and equipment management systems, can determine short-term and long-term equipment demand. By applying in a dynamic data-driven approach, the data could be dynamically and continuously updated. As such, the equipment usage prediction will be more reliable compared to the empirical method.

Data-Driven Equipment Combinations

Construction equipment is combined for certain operations in the construction project. For example, excavators are typically matched with trucks for earthmoving operations; thus, any calculations need to include the equipment type, haul distance, and whether the truck or the equipment is a constraint resource (Morley 2013; Liu 2014).

In construction engineering, both quantitative formulas and simulation methods are widely accepted and used by the practitioners for equipment combinations. Deriving quantitative formulas does take time; however, the investment can be easily justified by repeated use. Simulation models can also determine equipment combinations, but they are generally applied once or twice before needing modification. However, previous studies rarely investigate historical data of equipment combinations in the equipment management or project management systems.

In the future, when historical data of equipment combinations is readily available, data-driven, simulation-based analytics methods should be applied. The fleet combinations in the previous projects could be investigated to provide analytical support to the operation manager. It could further assist the fleet replacement plan, by converteing the empirical methods for planning into analytical methods capable of dynamically and continuously updating.

REFERENCES

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

Alshibani, A. and Moselhi, O. (2016). "Productivity based method for forecasting cost & time of earthmoving operations using sampling GPS data." *Journal of Information Technology in Construction (ITcon),* 21(3), 39-56.

American Society of Agricultural Engineers (ASAE). (2003). "Agricultural machinery management data." *ASAE D497.4*, American Society of Agriculture and Biological Engineers, St. Joseph, MI.

Amini, A.A., Mashayekhi, M., Ziari, H. and Nobakht, S. (2012). "Life cycle cost comparison of highways with perpetual and conventional pavements." *International Journal of Pavement Engineering*, 13(6), 553-568.

Aven, T. (2016). "Risk assessment and risk management: Review of recent advances on their foundation." *European Journal of Operational Research*, 253(1), 1-13.

Azar, E.R. and Kamat, V.R. (2017). "Earthmoving equipment automation: a review of technical advances and future outlook." *Journal of Information Technology in Construction (ITcon)*, 22(13), 247-265.

Bates, J., Rayner, A. and Custance, P. (1979). "Inflation and farm tractor replacement in the US: a simulation model." *American Journal of Agricultural Economics*, 61(2), 331.

Bayzid, S.M., Mohamed, Y. and Al-Hussein, M. (2016). "Prediction of maintenance cost for road construction equipment: a case study." *Canadian Journal of Civil Engineering,* 43(5), 480-492.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*, Springer, New York, NY.

Biller, B. and Gunes, C. (2010). "Introduction to simulation input modeling." *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, Maryland, USA

Blondel, V.D., Guillaume, J., Lambiotte, R. and Lefebvre, E. (2008). "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment,* 2008(10), P10008.

Borgatti, S.P. and Foster, P. C. (2003). "The network paradigm in organizational research: A review and typology." *Journal of Management,* 29(6), 991-1013.

Chae, S. and Yoshida, T. (2010). "Application of RFID technology to prevention of collision accident with heavy equipment." *Automation in Construction,* 19(3), 368-374.

Chan, A., Keoleian, G. and Gabler, E. (2008). "Evaluation of life-cycle cost analysis practices used by the Michigan Department of Transportation." *Journal of Transportation Engineering*, 134(6), 236-245.

Chinowsky, P. and Taylor, J.E., 2012. "Networks in engineering: an emerging approach to project organization studies." *Engineering Project Organization Journal*, 2(1-2), 15-26.

Choi, Y. and Nieto, A. (2011). "Optimal haulage routing of off-road dump trucks in construction and mining sites using Google Earth and a modified least-cost path algorithm." *Automation in Construction,* 20(7), 982-997.

Chung, T., Mohamed, Y. and AbouRizk, S. (2006). "Bayesian updating application into simulation in the NEST tunnel project." *Journal of Construction Engineering and Management*, 132(8), 882–894.

Cross, T.L. and Perry, G.M. (1995). "Depreciation patterns for agricultural machinery." *American Journal of Agricultural Economics*, 77(1), 194-204.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological),* 39(1), 1-22.

Drinkwater, R.W. and Hastings, N.A. (1967). "An economic replacement model." *Journal of the Operational Research Society*, 18(2), 121-138.

El-Rayes, K. and Moselhi, O. (2001). "Optimizing resource utilization for repetitive construction projects." *Journal of Construction Engineering and Management,* 127(1), 18-27.

Er, K.C., Fernando, S., AbouRizk, S. and Ruwanpura, J. (2000). "Selecting the best tunnel construction for the 1st stage of the SESS tunnel using simulation. " *7th Annual Construction Research Forum*, Edmonton, Canada.

Fan, H., AbouRizk, S., Kim, H. and Zaïane, O. (2008). "Assessing residual value of heavy construction equipment using predictive data mining model." *Journal of Construction Engineering and Management,* 22(3), 181-191.

Fan, H., Kim, H. and Zaïane, O.R. (2006). "Data warehousing for construction equipment management." *Canadian Journal of Civil Engineering,* 33(12), 1480-1489.

Forgy, E.W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21(1), 768-769.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). *Bayesian data analysis*, Chapman and Hall, London, UK.

Ghadam, P., Ravanshadnia, M. and Ramezani, S. (2012). "Determining Economic Life of Earth Moving Equipment by Using Life Cycle Cost Analysis: Case Study." *Proceedings of International Conference on Construction and Real Estate Management*, 2012, 1-6.

Gransberg, D.D. and O′Connor, E.P. (2015). *Major equipment life-cycle cost analysis*. Minnesota Department of Transportation, Research Services & Library. Minnesota Department of Transportation, St. Paul, MN

Hajjar, D. and AbouRizk, S. (1997) "AP2-Earth: A simulation-based system for the estimating and planning of earth moving operations." *Proceedings of the 1997 Winter Simulation Conference*, Atlanta, Georgia, USA.

Halpin, D.W. (2010). *Construction Management*. John Wiley & Sons, Hoboken, NJ.

Hansen, M.T., Mors, M.L. and Løvås, B. (2005). "Knowledge sharing in organizations: Multiple networks, multiple phases." *Academy of Management Journal,* 48(5), 776-793.

Han, J., Pei, J. and Kamber, M. (2011). *Data mining: Concepts and techniques.* Elsevier, Waltham, MA.

Hastings, W.K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1), 97-109.

Hendi, A. (2007). "Decision support system for equipment selection in construction projects." *Masters Abstracts International*, 46(3).

Ibrahim M. and Moselhi O. (2014). "Automated productivity assessment of earthmoving operations". *Journal of Information Technology in Construction (ITcon)*, 19, 169-184.

Ji, W. and AbouRizk, S.M. (2017). "Credible interval estimation for fraction nonconforming: Analytical and numerical solutions." *Automation in Construction*, 83(2017), 56-67.

Ji, W. and AbouRizk, S.M. (2018). "Simulation-Based Analytics for Quality Control Decision Support: Pipe Welding Case Study." *Journal of Computing in Civil Engineering*, 32(3), 05018002.

Kauffmann, P., Williams, R. and Hildreth, J. (2013). "Fleet Management Performance Monitoring." <https://connect.ncdot.gov/projects/research/RNAProjDocs/2012-07FinalReport.pdf> (February 11, 2019).

Kleiner, F. and Friedrich, H.E. (2017). “Maintenance and repair cost calculation and assessment of resale value for different alternative commercial vehicle powertrain technologies.” *Proceedings of the EVS30 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium*, European Association for Electromobility, Brussels, Belgium, 1-16.

Kuhl, M.E., Lada, E.K., Steiger, N.M., Wagner, M.A. and Wilson, J.R. (2009). "Introduction to modeling and generating probabilistic input processes for simulation." *Proceedings of the 2009 Winter Simulation Conference*, Austin, Texas, USA

Law, A.M. (2007). *Simulation Modeling and Analysis*. McGraw-Hill Education. New York City, NY.

Leemis, L.M. (2004). "Building credible input models." *Proceedings of the 2004 Winter Simulation Conference*, Piscataway, NJ.

Lei, Z. (2015). *Automated Simulation Model for Crane Motion Planning in Heavy Industrial Projects*. Doctoral Dissertation. University of Alberta, Edmonton, Canada.

Lin, C. and Tan, H. (2013). "Performance measurement in the public sector: Example of the building administration authorities in Taiwan." *Journal of Management in Engineering,* 30(1), 97-107.

Liu, C. (2014). "*Optimization of Temporary Haul Road Design and Earthmoving Job Planning based on Site Rough-grading Design*". M.Sc. Thesis, University of Alberta, Edmonton, AB

Liu, C., Ali, M. and AbouRizk, S. (2018). "Markov Chain-based Long-term Prediction of Equipment Usage hours from Historical Data". *CSCE 2018 Annual Conference in Fredericton*, Fredericton, NB

Lu, M., Chen, W., Shen, X., Lam, H.C. and Liu, J. (2007). "Positioning and tracking construction vehicles in highly dense urban areas and building construction sites. " *Automation in Construction*, 16(5), 647-656.

Lucko, G. (2003). *A statistical analysis and model of the residual value of different types of heavy construction equipment*, Doctoral Dissertation. Virginia Institute of Technology, Blacksburg, VA.

Lucko, G. Anderson-Cook, C.M. and Vorster, M.C. (2006). "Statistical considerations for predicting residual value of heavy equipment." *Journal of Construction Engineering and Management*, 132(7), 723-732.

Lucko, G., Vorster, M.C. and Anderson-Cook, C.M. (2007). "Unknown element of owning costs—Impact of residual value." *Journal of Construction Engineering and Management*, 133(1), 3-9.

Lucko, G. and Mitchell Jr., Z.W. (2010). "Quantitative research: Preparation of incongruous economic data sets for archival data analysis." *Journal of Construction Engineering and Management*, 136(1), 49-57.

Lucko, G. (2011). “Modeling the residual market value of construction equipment under changed economic conditions.” *Journal of Construction Engineering and Management*, 137(10), 806-816.

Mandapaka, V., Basheer, I., Sahasi, K., Ullidtz, P., Harvey, J.T. and Sivaneswaran, N. (2012). "Mechanistic-empirical and life-cycle cost analysis for optimizing flexible pavement maintenance and rehabilitation." *Journal of Transportation Engineering*, 138(5), 625-633.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). “Equation of state calculations by fast computing machines.” *The Journal of Chemical Physics*, 21(6), 1087-1092.

Mitchell Jr., Z.W. (1998). *A statistical analysis of construction equipment repair costs using field data and the cumulative cost model*, Doctoral Dissertation. Virginia Institute of Technology, Blacksburg, VA.

Mitchell, Z., Hildreth, J. and Vorster, M. (2011). "Using the cumulative cost model to forecast equipment repair costs: Two different methodologies." *Journal of Construction Engineering and Management.*, 137(10), 817-822.

Mithraratne, N. and Vale, B. (2004). "Life cycle analysis model for New Zealand houses." *Building and Environment*, 39(4), 483-492.

Morley, D., Lu, M. and AbouRizk, S. (2013) "Utilizing simulation derived quantitative formulas for accurate excavator hauler fleet selection." *2013 Winter Simulations Conference (WSC)*, 3018-3029.

Newman, M., Barabasi, A. and Watts, D.J. (2011). *The structure and dynamics of networks.* Princeton University Press, Princeton, NJ.

Peña-Mora, F., Han, S., Lee, S. and Park, M. (2008). "Strategic-operational construction management: Hybrid system dynamics and discrete event approach." *J. Constr. Eng. Manage.*, 134(9), 701–710.

Peurifoy, R., Ledbetter, W. and Schexnayder, C.J. (2002). *Construction, planning, equipment, and methods*, McGraw-Hill Education. New York City, NY.

Ponnaluru, S.S. (2009). *Empirical analysis of used construction equipment and auction house revenues*, Doctoral Dissertation. Washington State University, Pullman, WA.

Ponnaluru, S.S., Marsh, T.L. and Brady, M. (2012). "Spatial price analysis of used construction equipment: the case of excavators." *Construction Management and Economics,* 30(11), 981-994.

Pradhananga N. and Teizer J. (2015). "Cell-based construction site simulation model for earthmoving operations using real-time equipment location data." *Visualization in Engineering*, 3(12), 1-16.

Prasad Nepal, M. and Park, M. (2004). "Downtime model development for construction equipment management." *Engineering, Construction and Architectural Management,* 11(3), 199-210.

Prism Economics and Analysis and the Department of Civil Engineering, 2001, "*A Guide to Construction Cost Sources"*, University of Toronto

Priven, V. and Sacks, R. (2015). "Effects of the last planner system on social networks among construction trade crews." *Journal of Construction Engineering and Management.,* 141(6), 04015006.

Pryke, S. (2012). *Social network analysis in construction.* John Wiley & Sons, Hoboken, NJ.

Pryke, S.D. (2004). "Analysing construction project coalitions: exploring the application of social network analysis." *Constr. Manage. Econ.,* 22(8), 787-797.

Puri, V. and Martinez, J.C. (2012). "Modeling of simultaneously continuous and stochastic construction activities for simulation." *Journal of Construction Engineering and Management*, 139(8), 1037-1045.

R Core Team (2019). "R: A language and environment for statistical computing." <https://www.R-project.org/> (December 25, 2019).

Raiffa, H. and Schlaifer, R. (2000). *Applied statistical decision theory.* Wiley, Hoboken, NJ.

Robert C.P. and Casella G. (1999). *Monte Carlo Statistical Methods*. Springer, New York, NY

Santos, J. and Ferreira, A. (2013). "Life-cycle cost analysis system for pavement management at project level." *International Journal of Pavement Engineering*, 14(1), 71-84.

Samee, K. and Pongpeng, J. (2016). "Structural equation model for construction equipment management affecting project and corporate performance." *KSCE Journal of Civil Engineering,* 20(5), 1642-1656.

Sandhu, M. and Helo, P. (2006). "A network approach to project business analysis." *Engineering, Construction and Architectural Management,* 13(6), 600-615.

SAP SE (2018). *Proven, time-tested on-premise ERP.* Toronto, Ontario, Canada, available at: https://www.sap.com/canada/products/enterprise-management-erp.html (accessed May 18, 2018).

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

Tortoriello, M., Reagans, R. and McEvily, B. (2012). "Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units." *Organization Science,* 23(4), 1024-1039.

Vorster, M. C. (2009). *Construction equipment economics.* Pen, Christiansburg, VA.

Vorster, M.C. (1980). *A systems approach to the management of civil engineering construction equipment*. Doctoral Dissertation. Stellenbosch University, Stellenbosch, South Africa.

Wang, H., Zhang, J., Chau, K. and Anson, M. (2004). "4D dynamic management for construction planning and resource utilization." *Automation in Construction.,* 13(5), 575-589.

Wang, L. and Zhang, X. (2018). “Bayesian analytics for estimating risk probability in PPP waste-to-energy projects.” *Journal of Management in Engineering*, 34(6), 04018047.

Wang, Q. and Taylor, J.E., 2015. "Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms." *Journal of Computing in Civil Engineering*, 30(2), 04015004.

Werner, M., Ji, W. and AbouRizk, S. (2018). “Improving tunneling simulation using Bayesian updating and hidden Markov chains.” *Proceedings of the 2018 Winter Simulation Conference*, IEEE, Piscataway, NJ, 3930-3940.

Yip, H., Fan, H. and Chiang, Y. (2014). "Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models." *Automation in Construction,* 38 30-38.

Zhang, H. and Li, H. (2004). "Simulation-based optimization for dynamic resource allocation." *Automation in Construction,* 13(3), 409-420.

Zhang, L. and Ashuri, B. (2018). "BIM log mining: Discovering social networks." *Automation in Construction,* 91 31-43.

Zhang, L., Ekyalimpa, R., Hague, S., Werner, M. and AbouRizk, S. (2015). “Updating geological conditions using Bayes theorem and Markov chain.” *Proceedings of the 2015 Winter Simulation Conference,* IEEE, Piscataway, NJ, 3367-3378.

Zhang, S., Du, C., Sa, W., Wang, C. and Wang, G. (2013). “Bayesian-based hybrid simulation approach to project completion forecasting for underground construction.” *Journal of Construction Engineering and Management*, 140(1), 04013031.

Zhao, T. and Fu, C.C. (2006). “Infrastructure development and expansion under uncertainty: A risk-preference-based lattice approach.” *Journal of Construction Engineering and Management*, 132(6), 620-625.

Zheng, X., Le, Y., Chan, A. P., Hu, Y. and Li, Y. (2016). "Review of the application of social network analysis (SNA) in construction project management research." *International Journal of Project Management,* 34(7), 1214-1225.

Zong, Y. (2017). *Maintenance cost and residual value prediction of heavy construction equipment*, M.Sc. Thesis. University of Alberta, Edmonton, Canada.

APPENDIX A. Bargain Likelihood Achievement by Using Monte Carlo Simulation

In the Chapter 2, the bargain likelihood is represented by a mean value since it can be easily understood and used by practitioners, but it can also be defined as a distribution, which is the product of two posterior distributions. As such, the distribution of bargain likelihood cannot be directly achieved through the proposed Eq. 2.8.

Instead, a Monte Carlo simulation is used to obtain the distribution of bargain likelihood through iterations. In each iteration, two random numbers are generated from the posterior distributions (i.e., beta distribution for bargain percentage, and normal distribution for transaction quantity, respectively, Table 2.3) and, one randomized value of bargain likelihood can be achieved as the product of two random numbers. Given the posterior distributions of bargain percentage and transaction quantity, randomized values of bargain likelihood can be collected to form the distribution of bargain likelihood, which allows for quantifying the uncertainty of bargain likelihood along with its mean value.

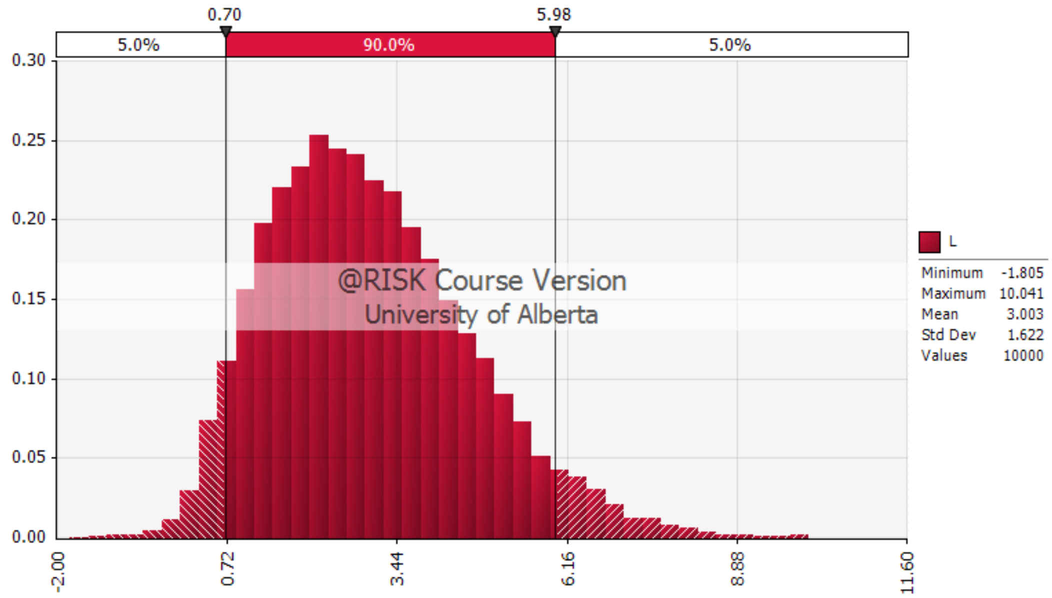
The following illustrative example explains in detail the achievement of bargain likelihood through Monte Carlo simulation. Assuming the poster distributions are Beta (3,2) and Norm (5,2), 10,000 random numbers are generated from each distribution. For the first 10 iterations, the normalized numbers are listed as the following table.

Iteration #	Random Number for Bargain Percentage	Random Number for Transaction Quantity
1	0.573	5.353
2	0.401	4.326
3	0.277	3.736
4	0.494	3.139
5	0.945	3.028
6	0.881	5.251
7	0.449	6.394
8	0.773	4.781
9	0.912	5.930
10	0.730	7.458
…	…	…

The randomized value of bargain likelihood can be achieved as the following table.

Iteration #	Random Number for Bargain Percentage (B)	Random Number for Transaction Quantity (Q)	Randomized Value of Bargain Likelihood (L=B×Q)
1	0.573	5.353	3.066
2	0.401	4.326	1.734
3	0.277	3.736	1.034
4	0.494	3.139	1.551
5	0.945	3.028	2.860
6	0.881	5.251	4.625
7	0.449	6.394	2.871
8	0.773	4.781	3.695
9	0.912	5.930	5.410
10	0.730	7.458	5.443
…	…	…	…

After 10,000 iterations, 10,000 randomized value of bargain likelihood are achieved, and the histogram (or distribution) of bargain likelihood is demonstrated as the following figure.



According to the figure, 90% confidence interval of the bargain likelihood is [0.70, 5.98] and practitioner can quantify the uncertainty of bargain likelihood. Notably, the mean of 10,000 randomized value is 3.003 while the mean value of bargain likelihood calculated directly through Eq.2.8 is 3 ($\frac{3}{3+2}\times 5=3$).

APPENDIX B. Manual Calculation of DDI

In the illustrative example of Chapter 4, the values for DDI are calculated using Eq. 4.3 and Eq. 4.4, as 0 and 0.7125 for Models A and B, respectively. The detailed calculations are explained as follows:

Calculation of DDI for Model A:

Edge ID	Node i	Node j	(1) w_{ij}	(2) d_{ij}	(3) d_{ij}^*	(4) $(d_{ij}^* - d_{ij})$	(5) $(d_{ij}^* - d_{ij}) \times w_{ij}$	(6) $d_{ij}^* \times w_{ij}$
1	S	1	10	10	10	0	0	100
2	S	2	10	10	10	0	0	100
3	S	3	10	10	10	0	0	100
4	S	4	10	10	10	0	0	100
5	1	2	0	1	20	19	0	0
6	2	3	0	1	20	19	0	0
7	3	4	0	1	20	19	0	0

* Based on the Equation 4.3, $d_{ij}^* = d_{is} + d_{sj}$. If node i or j is the shop (S), $d_{ij}^* = d_{ij}$.

Therefore, the numerator of Equation 4.4 is the sum of column (5), which is 0, and the dominator of Equation 4.4 is the sum of column (6) which is 400. Thus, the DDI is calculated as 0 (=0/400) for Model A.

The calculation of DDI for Model B:

Edge ID	Node i	Node j	(1) w_{ij}	(2) d_{ij}	(3) d_{ij}^*	(4) $(d_{ij}^* - d_{ij})$	(5) $(d_{ij}^* - d_{ij}) \times w_{ij}$	(6) $d_{ij}^* \times w_{ij}$
1	S	1	5	10	10	0	0	50
2	S	2	0	10	10	0	0	0
3	S	3	0	10	10	0	0	0
4	S	4	5	10	10	0	0	50
5	1	2	5	1	20	19	95	100
6	2	3	5	1	20	19	95	100
7	3	4	5	1	20	19	95	100

*Based on the Equation 4.3, $d_{ij}^* = d_{is} + d_{sj}$. If node i or j is the shop (S), $d_{ij}^* = d_{ij}$.

Therefore, the numerator of Equation 4.4 is the sum of column (5) which is 285 and the dominator of Equation 4.4 is the sum of column (6) which is 400. Thus, the DDI is calculated as 0.7125 (=285/400) for Model B.

Clustering coefficient can be calculated for any node in the network according to the social network analysis. Similarly, DDI can be calculated for any node in the network. In the DDI calculations, the node we are studying is assumed to be the equipment storage center or equipment shop. The nodes not representing the equipment shop in the real practice will achieve a relatively low value in DDI, which is not helpful in evaluating the equipment management performance. Thus, our study focuses on the DDI calculation of nodes representing only the equipment shop.

APPENDIX C. R Code 1 – K-Means Clustering

```
##### K-Means Clustering
#' maxk is the limit of clustering number, which is 10 in this case
Input_Data=Raw_Data[,-1]
maxk=10
dif=(1:maxk)
for (k in 2:maxk){
    km.out = kmeans(Input, k, nstart=20)
    km.out$cluster
    a=(1:(k-1))
    b=a
    j=1
    for(i in 1:(length(km.out$cluster)[1]-1)){
        n=km.out$cluster[i]
        m=km.out$cluster[i+1]
        if (n!=m){
            a[j]=i
            b[j]=i+1
            j=j+1
        }
    }
  dif[k]=0
  for (i in 1:(k-1)){ dif[k]= dif[k]+( Input_Data [a[i],2]- Input_Data[b[i],2])^2/(k-1) }
  maxdif=max(dif)
```

APPENDIX D. R Code 2 – EM Algorithm

```
###Initialize
SP.kmeans <- kmeans(SP, 2)
SP.kmeans.cluster <- SP.kmeans$cluster
SP.df <- data_frame(x = SP, cluster = SP.kmeans.cluster)

SP.df %>%
  mutate(num = row_number()) %>%
  ggplot(aes(y = num, x = x, color = factor(cluster))) +
  geom_point() +
  ylab("Values") +
  ylab("Data Point Number") +
  scale_color_discrete(name = "Cluster") +
  ggtitle("K-means Clustering")

SP.summary.df <- SP.df %>%
  group_by(cluster) %>%
  summarize(mu = mean(x), variance = var(x), std = sd(x), size = n())

SP.summary.df <- SP.summary.df %>%
  mutate(alpha = size / sum(size))
SP.summary.df <- SP.df %>%
group_by(cluster) %>%
summarize(mu = mean(x), variance = var(x), std = sd(x), size = n())
SP.summary.df = SP.summary.df[,-5]
```

```
#' Expectation Step of the EM Algorithm

#' Calculate the posterior probabilities (soft labels) that each component has to each data point.

#'#' @param sd.vector Vector containing the standard deviations of each component

#' @param sd.vector Vector containing the mean of each component

#' @param alpha.vector Vector containing the mixing weights of each component

#' @return Named list containing the loglik and posterior.df

e_step <- function(x, mu.vector, sd.vector, alpha.vector) {

  comp1.prod <- dnorm(x, mu.vector[1], sd.vector[1]) * alpha.vector[1]

  comp2.prod <- dnorm(x, mu.vector[2], sd.vector[2]) * alpha.vector[2]

  sum.of.comps <- comp1.prod + comp2.prod

  comp1.post <- comp1.prod / sum.of.comps

  comp2.post <- comp2.prod / sum.of.comps

  sum.of.comps.ln <- log(sum.of.comps, base = exp(1))

  sum.of.comps.ln.sum <- sum(sum.of.comps.ln)

  list("loglik" = sum.of.comps.ln.sum,

       "posterior.df" = cbind(comp1.post, comp2.post))

}


#' Maximization Step of the EM Algorithm

#' Update the Component Parameters

#' @param x Input data.

#' @param posterior.df Posterior probability data.frame.

#' @return Named list containing the mean (mu), variance (var), and mixing

#'   weights (alpha) for each component.

m_step <- function(x, posterior.df) {
```

```
  comp1.n <- sum(posterior.df[, 1])
  comp2.n <- sum(posterior.df[, 2])
  comp1.mu <- 1/comp1.n * sum(posterior.df[, 1] * x)
  comp2.mu <- 1/comp2.n * sum(posterior.df[, 2] * x)
  comp1.var <- sum(posterior.df[, 1] * (x - comp1.mu)^2) * 1/comp1.n
  comp2.var <- sum(posterior.df[, 2] * (x - comp2.mu)^2) * 1/comp2.n
  comp1.alpha <- comp1.n / length(x)
  comp2.alpha <- comp2.n / length(x)
  list("mu" = c(comp1.mu, comp2.mu),
       "var" = c(comp1.var, comp2.var),
       "alpha" = c(comp1.alpha, comp2.alpha))
}

######################
for (i in 1:50) {
  if (i == 1) {
    # Initialization
    e.step <- e_step(SP, SP.summary.df[["mu"]],
SP.summary.df[["std"]],SP.summary.df[["alpha"]])
    m.step <- m_step(SP, e.step[["posterior.df"]])
    cur.loglik <- e.step[["loglik"]]
    loglik.vector <- e.step[["loglik"]]
  } else {
    # Repeat E and M steps till convergence
    e.step <- e_step(SP, m.step[["mu"]], sqrt(m.step[["var"]]), m.step[["alpha"]])
    m.step <- m_step(SP, e.step[["posterior.df"]])
```

```
    loglik.vector <- c(loglik.vector, e.step[["loglik"]])
    loglik.diff <- abs((cur.loglik - e.step[["loglik"]]))
    if(loglik.diff < 1e-6) {
      break
    } else {
      cur.loglik <- e.step[["loglik"]]
    }
  }
}
loglik.vector

#######################
#' Plot a Mixture Component
#' @param x Input ata.
#' @param mu Mean of component.
#' @param sigma Standard of component.
#' @param lam Mixture weight of component.
plot_mix_comps <- function(x, mu, sigma, lam) {
  lam * dnorm(x, mu, sigma)
}
```

APPENDIX E. R Code 3 – Bayesian Updating

```
##MH Function
   # likelihood: a function of the parameter vector
   # prior: a function of the parameter vector
   # proposal: a function of the current value of the parameter
   # vector, returning a proposed parameter vector
   # pars0: a vector of initial values of the parameters
   # m: number of samples, defaults to 10^5
metropolis = function(likelihood,prior,proposal,pars0,m=1e4){
   post=function(pars) likelihood(pars)*prior(pars)
   pars=matrix(0,m,length(pars0))
   accepted<-rejected<-0
   pars.current<-pars0
   for (i in 1:m) {
         pars.proposed=proposal(pars.current)
         a=post(pars.proposed)/post(pars.current) ##acceptance ratio
         if(a>=1 || runif(1)<=a) {
         pars[i,]=pars.proposed
         pars.current=pars.proposed
         accepted=accepted+1
      }
      else{
         pars[i,]=pars.current
         rejected=rejected+1
      }
```

```
    }
    if(length(pars0) ==1) pars<-as.vector(pars)
    result=list(samples=pars, accepted=accepted, rejected=rejected, thinned=FALSE)
    class(result)="metropolis"
    result
}

##Summary MH Results
print.metropolis<-function(x, ...) {
    cat("number of samples: ", with (x, if(is.matrix(samples)) nrow(samples) else
length(samples)))
    cat("\nnumber of parameters: ", with (x, if(is.matrix(samples)) ncol(samples) else 1))
    if (x$thinned) {cat("\nPrior to thinning:") }
    cat("\nnumber of proposals accepted: " , x$accepted)
    cat("\nnumber of proposals rejected: " , x$rejected)
    cat("\npercentage of proposals accepted: ", with (x,
round(100*accepted/(accepted+rejected),2)),"\n")
    if (is.matrix(x$samples)) {
        cat("\nestimated posterior medians")
        print(apply(x$samples, 2, median))
    }
    else cat("\nestimated posterior median: ", median(x$samples))
    invisible(x)
}

##Plot MH Results
```

```
plot.metropolis <- function(x, ...) {
   n.par <- if(is.matrix(x$samples)) ncol(x$samples) else 1
   if (n.par==1) acf(x$samples, main="", ...)
   else{
      save.mfrow = par(mfrow = n2mfrow(n.par))
      on.exit(options(save.mfrow))
      for (j in 1:npar) {
         acf(x$samples[,j],main=paste("parameter", j), ...)
      }
   }
}

#Thin Method
thin<-function(object, ...) {UseMethod("thin")}
thin.metropolis<-function(object, by, ...) {
   #by: every by-th sample in object is retained
   samples<-object$samples
   object$samples <-if(is.matrix(samples)) {samples[seq(1, nrow(samples), by=by), ]}
   else    samples[seq(1, length(samples), by=by)]
   object$thinned <-TRUE
   object
}
```