CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK CONSTRUCTION IN IOWA USING SPATIAL STATISTICAL MODELING

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Martha Sebsibe Hitsanu

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Construction Management and Engineering

March 2022

Fargo, North Dakota

North Dakota State University
Graduate School

Title

CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK CONSTRUCTION IN IOWA USING SPATIAL STATISTICAL MODELING

By

Martha Sebsibe Hitsanu

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Eric Asa
Chair

Dr. Majura Selekwa

Dr. Stephanie S. Day

Approved:

Apr 12, 2022 — Date

Dr. Xuefeng (Michael) Chu — Department Chair

ABSTRACT

Estimating cost at the conceptual phase of highway earthwork construction projects is challenging owing to the limited amount of information present at the initial stages. Methods that have applied in the past to perform such estimates are regression, mathematical equations, artificial intelligence and others. The objective of this research is to use two spatial statistical methods – indicator kriging and empirical Bayesian kriging – together with exponential, whittle and k-Bessel semi-variograms to build an appropriate model to predict highway earthwork construction cost at the conceptual stage. Iowa Department of Transportation provided the data for the study. The cross-validation results were used to compare the performance of the kriging algorithms. It was observed that the exponential semi-variogram showed better performance for both types of kriging when they were analyzed separately. However, indicator kriging results were more accurate than empirical Bayesian kriging when they were compared for the same type of semi-variogram.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my academic advisor Dr. Eric Asa for accepting me as his student and the continuous support he provided throughout my study. I would also like to thank the committee members for their time and guidance. I would like to thank my mom, family and friends who supported, motivated, and prayed for me without ceasing. I would also like to thank the department of construction management and engineering at North Dakota State University for supporting me financially.

DEDICATION

I dedicate this work to God (the Perfecter and Finisher). I am eternally grateful for His guidance and support thought out my life.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS iv

DEDICATION v

1. INTRODUCTION 1

1.1. Background 1

1.2. Problem Statement 1

1.3. Research Methodology 5

1.4. Research Contribution 6

1.5. References 8

2. A SYSTEMATIC LITERATURE REVIEW ON CONCEPTUAL COST ESTIMATIONS OF HIGHWAY EARTHWORK CONSTRUCTION 10

2.1. Background 10

2.2. Previous Studies 11

2.3. Research Questions 14

2.4. Search Process 15

2.5. Cost Estimation Methods Used for the Conceptual Stage of Highway Projects 16

2.6. Factors Affecting Highway Earthwork Construction Cost 20

2.7. Discussion of Results 22

2.8. Conclusion 23

2.9. References 24

3. EXPLORATORY DATA ANALYSIS OF HIGHWAY EARTHWORK BID ITEMS 27

3.1. Introduction 27

3.2. Data Collection and Preparation 31

3.3. Outlier Detection and Removal 33

3.4. Normality Test and Transformation 35

3.5. Conclusion 43

3.6. References 43

4. CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK UNIT PRICE USING EMPIRICAL BAYESIAN KRIGING 46

4.1. Introduction 46

4.2. Semi Variogram Estimation 52

4.3. Comparison of Models by Cross Validation 54

4.4. Conclusion 57

4.5. References 58

5. CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK UNIT PRICES: INDICATOR KRIGING APPROACH 60

5.1. Introduction 60

5.2. Materials and Methods 63

5.3. Semi Variogram Modeling 70

5.4. Cross Validation 74

5.5. Conclusion 79

5.6. References 80

6. COMPARISON OF INDICATOR KRIGING AND EMPIRICAL BAYESIAN KRIGING FOR CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK BID ITEMS 83

6.1. Introduction 83

6.2. Material and Methodology 85

6.3. Cross Validation Results 90

6.4. Conclusion 96

6.5. References 97

7. CONCLUSION AND RECOMMENDATION 99

LIST OF FIGURES

Figure		Page
1.1:	Research methodology flow chart	5
3.1:	2015 embankment in place bid item data & summary before outlier removal	32
3.2:	2015 embankment in place bid items & summary after outlier removal	34
3.3:	Outlier detection using Grubbs	34
3.4:	Histogram and QQ plot of skewed data set before log transformation	36
3.5:	Histogram and QQ plot of selected bid items after log transformation	37
4.1:	Empirical Bayesian kriging approach in developing price model	51
4.2:	Semi variogram models of sample bid items-empirical Bayesian kriging	53
4.3:	Cross-validation results for sample cost data-empirical Bayesian kriging	54
4.4:	Embankment 2019 price map	56
4.5:	Excavation roadway and borrow 2019 price map	56
4.6:	Excavation roadway and borrow 2018 price map	57
5.1:	Indicator kriging approach in developing price model	69
5.2:	K-Bessel semi-variograms before and after optimization-indicator kriging.	71
5.3:	Exponential and k-Bessel semi variograms for sample bid data-indicator kriging	72
5.4:	Cross-validation results for sample cost data-indicator kriging	75
5.5:	Embankment 2019 price map	78
5.6:	Excavation roadway and borrow 2018 price map	78
5.7:	Excavation roadway and borrow 2019 price map	79
6.1:	Research methodology	89
6.2:	Cross-validation result for embankment-comparison	91
6.3:	Cross-validation result for excavation boulder-comparison	92
6.4:	Cross-validation result for excavation roadway and borrow-comparison	93

1. INTRODUCTION

1.1. Background

Conceptual cost estimation is a kind of cost estimate that is prepared when the design for the project is 30% complete (Asmar et al. 2011). Conceptual cost estimation is key to the success in construction projects in which the project is said to be successful if the cost estimate is done accurately (Roxas et al. 2019). An accurate cost estimate will help construction project managers make good financial decisions (Sodikov 2005). When cost is accurately estimated, both the contractor and the client can secure good revenue and can have better control of project resources/funds (Roxas et al. 2019). As the project nears completion, the accuracy level of the cost estimation increases as more project information becomes available (Sodikov 2005).

Earthwork is a crucial component of highway construction and it is considered one of the most complicated processes in the construction of a highway project (Bogenberger et al. 2015). Earthwork encompasses drilling, blasting, moving, filling, and compacting vast amounts of granular materials (Bogenberger et al. 2015). Earthwork is costly and entails significant social and environmental impacts on the area surrounding the construction site (Bogenberger et al. 2015). Thus, there is a need for a better method to estimate the cost of a highway earthwork project at the conceptual phase.

1.2. Problem Statement

Earthwork accounts for the highest share of overall highway construction cost, as it involves the ownership and operating cost of heavy equipment used in the construction process (Lewis and Hajji 2012). Earthwork is the most extensive and tedious part when compared to other highway construction bid items, and it involves excavating, transporting, spreading, and compacting large quantities of soils (Bogenberger et al. 2015). Similarly, Uhlik (1984) stated that earthwork construction involves the greatest amount of construction work on a given project and it can bring a significant risk to the project if estimated inaccurately.

Crucial factors that affect the reliability of a construction project estimate include price increases, interruptions, and cost overruns (Roxas et al. 2019). It is particularly challenging to prepare an estimate at the conceptual phase because of a lack of sufficient project data (Asmar et al. 2011). In addition, there is a limited roadway cost database and limited cost estimating methods (Sodikov 2005). The precision of estimating construction cost at the beginning of a project is affected by inadequate data and unpredicted factors (Roxas et al. 2019).

Highway construction is a special type of construction whose price heavily depends on the location (Le et al. 2019). It is considered challenging for state highway agencies to perform an accurate estimation of the cost of a highway project at the conceptual phase (Asmar et al. 2011).

Several methods have been used to perform conceptual estimation of highway cost. Sodikov (2005) discovered what seemed like a more appropriate method for early estimation by using the artificial neural network (ANN) method. Similarly, Roxas et al. (2019) proposed a study that used an ANN for estimating road project costs but still has limitations associated with the scarcity of data. Asmar et al. (2011) used a method that closely resembles the program evaluation and review technique (PERT) which attached factors to bring more confidence into the estimate. The proposed method used historical bid data for those items whose estimate can be known in the earlier phase of the project and employs contingency factors for the other parts of the project using percentages. Although this method used a unique approach to examine historical data, it is stated that there was a limit in the analysis because of the scarcity of contingency item data. According to the study conducted by Zitta et al. (2019), earthwork cost estimation was performed using computer software programming and a mass haul diagram. The limitation stated by the study was that the software needed to be improved and be opened to others to verify the results. Similarly, in a study conducted in Nepal by Sthapit and Mori (1994), an equation was used to predict the cost with hill and slope factors as the dependent variables to estimate the cost of highway earthwork projects. Although the study was successful on a hundred-meter length road, it was stated that the validity of other road segments should be tested.

Different researchers have tried to tackle the issue, yet the problem of estimating construction cost at the early stage persists. Only a limited number of studies have been conducted for cost estimation of actual earthwork projects.

The objective of this research is to use two spatial statistical methods – indicator kriging and empirical Bayesian kriging – together with exponential, whittle and k-Bessel semi-variograms to build an appropriate model to predict highway earthwork construction cost at the conceptual stage and then compare them to choose which model gives the best result. A systematic literature review of cost estimation of highway earthworks will be carried out. The research gap will be addressed by applying two methods of kriging (indicator kriging and empirical Bayesian kriging), together with different kinds of semi-variograms. Geographic information system (GIS) will be used as the software medium to implement the geostatistical analysis. Data from the Iowa State Department of Transportation will be used. Outliers will be detected and removed, and exploratory data analysis (EDA) will be applied to the data to study the trend and perform transformations. Afterward, the kriging methods in combination with different semi-variograms will be applied. Cross-validation will then be used to check the model for errors and to ensure accuracy. The diagrammatic representation of the research flow is shown in Figure. 1

The research will address the following questions:

RQ 1: What methods are used for the conceptual cost estimation of highway earthwork construction projects?

RQ 2: What are the factors affecting highway construction cost estimates

RQ 3: Which semi-variogram combined with empirical Bayesian kriging gives better accurate conceptual cost estimation of highway earthwork? How are the models validated?

RQ 4: Which semi-variogram combined with indicator kriging gives better accurate conceptual cost estimation of highway earthwork? How are the models validated?

RQ 5: From indicator and empirical Bayesian kriging, for the same kind of semi-variograms, which kriging type gives better estimating accuracy? How are the models validated?

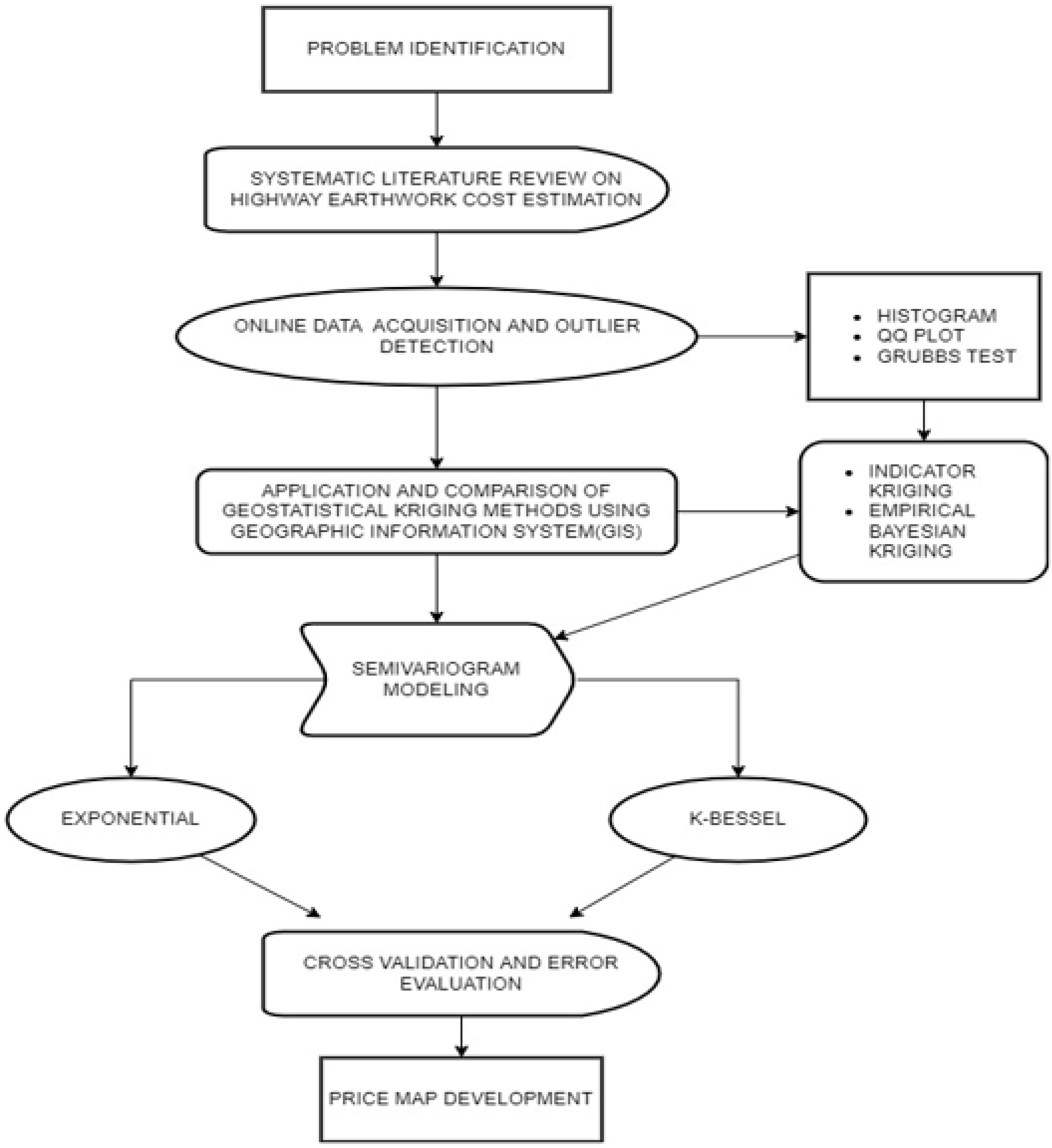


Fig. 1.1: Research methodology flow chart

1.3. Research Methodology

To achieve the objective of this research, the following research methodology is presented. Figure 1 describes the six chapters that together describe the research conducted:

This research was conducted to identify and evaluate the performance of two kriging methods to choose the one that gives the most accurate result. The earthwork cost data used for the study was acquired from the Iowa Department of Technology (DOT). Iowa DOT was chosen as a source of data for the study, because of the availability of the vast amount of data the department holds as compared to other departments of transportation. Bid Items of embankment, excavation boulder, and excavation roadway and borrow for the years 2015 to 2020 were chosen because of their availability. Visual Inspection was done on the data to clean abnormal data points. Then exploratory data analysis was conducted. Outliers were removed using histogram and QQ plot by careful investigation of points that were far away from the rest of the data. The Grubbs test in Minitab software was also implemented to detect and remove outliers. Two types of kriging algorithms: empirical Bayesian kriging and indicator kriging as combined with different types of semi-variograms (exponential, whittle and k-Bessel semi-variograms) were used, to evaluate the performance of the models and choose the more accurate model to predict highway earthwork cost at the conceptual stage. The analysis was performed for each of the three-unit items of work (embankment, excavation boulder, and excavation roadway and borrow) for each year of 2015 to 2020. Cross-validation was conducted on the two kriging types separately to identify which semi-variogram gives the less error value. In the last chapter, for the same kinds of semi-variograms, the performance of the two kriging types was tested by cross-validation. Cross-validation was implemented to validate all results and test the performance of the models. In the end price map was developed for each model.

1.4. Research Contribution

The goal of this research is to use two spatial statistical methods – indicator kriging and empirical Bayesian kriging – together with exponential, whittle and k-Bessel semi-variograms to compare and evaluate the performance of the two kriging methods and find an appropriate model to predict highway earthwork construction cost at the conceptual stage. Highway bid data will be collected from the Iowa Department of Transportation to build the model. GIS will be used as a medium to conduct the research. By accurately estimating the cost at the conceptual stage, this study will help in better controlling the construction budget. It will benefit both owners and construction stakeholders: architects, engineers, and contractors. Presently, there is a scarcity of information on cost estimation of highway earthwork projects. This study will contribute to the existing knowledge, the most appropriate method that better assists the conceptual cost estimation of highway earthwork projects by testing and evaluating the performance of two kriging methods, indicator kriging and empirical Bayesian kriging.

The research is comprised of seven chapters. Chapter one, which is the introduction chapter starts by giving the background information on highway earthwork projects and conceptual cost estimation. This chapter presents the research questions that are going to be addressed. Chapter two is the systematic literature review chapter. It assesses the studies currently performed and their limitations. It will search for articles to find answers to research questions one and two. Chapter three is the exploratory data analysis chapter. It will explore the behavior and trend of data by detecting outliers and performing normal distribution to assist the analysis of further chapters for the prediction of conceptual cost. Three highway earthwork cost bid items: embankment, excavation boulder, and excavation roadway and borrow will be extracted from the Iowa Department of Transportation for the years 2015 to 2020. The inflation factor will be applied to bring the values to 2021. Chapter four is conceptual cost estimation of highway earthwork unit price using empirical Bayesian kriging. In this chapter, empirical Bayesian kriging will be combined with exponential, whittle, and k-Bessel semi-variograms to choose which combination will perform better to predict the cost accurately. Cross-validation will be used for choosing the model, in which error values will be calculated and the model with less error will be chosen. This chapter answers research question three. Chapter five is conceptual cost estimation of highway earthwork unit price using indicator kriging. In this chapter indicator kriging will be combined with exponential and k Bessel semi-variograms to choose which combination gives a better result to predict the cost accurately. Cross-validation will be implemented to choose the best combination by assessing error values. This chapter will address research question four. Chapter six is the comparison of empirical Bayesian kriging and indicator kriging for the conceptual cost estimation of highway earthwork projects. In this chapter, both indicator kriging and empirical Bayesian kriging will be combined with exponential and k Bessel semi-variograms to compare and choose the one giving the best result with less error to predict cost by using cross-validation. This chapter answers research question five. And, the final chapter seven is the conclusion and recommendation chapter that concludes all the previous chapters by pointing out the findings and putting forwards recommendations that will open doors for future studies

1.5. References

Asmar, M. E., Hanna, A. S., & Whited, G. C. (2011). New approach to developing conceptual cost estimates for highway projects. Journal of construction engineering and management, 137(11), 942-949

Bogenberger, C., Dell’ Amico, M., Fuellerer, G., Hoefinger, G., Iori, M., Novellani, S., & Panicucci, B. (2015). Two-phase earthwork optimization model for highway construction. Journal of construction engineering and management, 141(6), 05015003.

Le, C., Le, T., Jeong, H. D., & Lee, E. B. (2019). Geographic information system–based framework for estimating and visualizing unit prices of highway work items. Journal of Construction Engineering and Management, 145(8), 04019044.

Lewis, P., & Hajji, A. (2012). Estimating the economic, energy, and environmental impact of earthwork activities. In Construction Research Congress 2012: Construction Challenges in a Flat World (pp. 1770-1779).

Roxas, C. L. C., Roxas, N. R., Cristobal, J., Hao, S. E., Rabino, R. M., & Revalde, F. (2019). Modeling road construction project cost in the Philippines using the artificial neural network approach. In 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (pp. 1-5). IEEE.

Sodikov, J. (2005). Cost estimation of highway projects in developing countries: artificial neural network approach. Journal of the Eastern Asia Society for Transportation Studies, 6, 1036-1047.

Sthapit, N., & Mori, H. (1994). Model to estimate highway earthwork cost in Nepal. Journal of transportation engineering, 120(3), 498-504.

Uhlik III, F. T. (1984). Optimizing earthwork estimating for highway construction. The Pennsylvania State University.

Zitta, N., Aliyu, M. R., Peter, E. U., & Ayuba, Y. (2019). Estimating Cost of Earthwork Using Computerized Visual Display Approach. ATBU Journal of Science, Technology and Education, 7(1), 138-148.

2. A SYSTEMATIC LITERATURE REVIEW ON CONCEPTUAL COST ESTIMATIONS OF HIGHWAY EARTHWORK CONSTRUCTION

2.1. Background

Cost estimation of highway earthwork at the early phase is a challenging process. Right after the layout for a given road project is set, earthwork construction is the first stage of the highway construction processes (Bogenberger et al. 2015). Once the path for a certain route is set out, the road design will pass through two steps: the first step is the selection of grade line by balancing the cut and the fill quantity and the second step is setting of the cross-section and calculation of the volume of the earthwork (Easa 1988). Highway earthwork construction involves the process of moving earth materials from high-grade to low-grade positions and involves deciding which areas to cut and fill for volume calculation (Bogenberger et al. 2015). Cutting and filling of earth materials is how the natural road level is brought to the desired road level (Zitta et al. 2019). According to Bogenberger et al. (2015), the main works involved in an earthwork activity are digging, transporting, filling, and compacting of earth materials. The study recommends this work should be done with extra caution. The researchers emphasized that earthwork construction carries the highest proportion of construction work, which makes it become the key component. Road earthwork construction is expensive and complicated (Bogenberger et al. 2015). It also has an extensive impact on the environment, which is why it is highly regulated (Bogenberger et al. 2015). The amount of money allocated to earthwork construction varies from one construction site to another (Kim et al. 2018).

Earthwork uses various kinds of machines/equipment, which can come in a single or a group set (Hola and Schabowicz 2010). In addition to its high cost, earthwork often takes an enormous amount of time for its completion (Parente et al. 2014). Since earthwork is so integral and takes the highest proportion of highway construction bid items, the planning should be done with extra care (Jayawardane and Harris 1990). Yet, there is still a challenge to accurately estimate the cost before construction (Zitta et al. 2019). Bogenberger et al. (2015) also emphasized the challenge in trying to find the minimal cost associated with the transportation of earth from cut to fill locations. Given the significance of highway earthwork construction in terms of volume and cost, there is a need for an improved conceptual cost estimation method.

The study addressed the following research questions:

RQ 1: What are the methods which have been used in the conceptual cost estimation of highway earthwork construction and their limitations?

RQ 2: What are the factors affecting highway earthwork costs?

2.2. Previous Studies

When any new state highway project is initiated, the main aim is completing it within budget since cost overruns ruin the reputation of builders (Zhang 2017). Conceptual cost estimation is defined as a kind of estimation which predicts the future project cost when the design is in the initial stage of completion (Mahamid 2013). Asmar et al. (2011) defined conceptual cost estimate as to the kind of estimate that is performed when the design is almost 30% complete and when only a limited amount of information exists about the project. The study also highlights this early estimation of cost not only includes the direct construction costs, but also the indirect construction costs, which include right of way and other indirect costs. Accurate early estimation is important because all future estimates will be based on the preliminary estimation, whether it was done correctly or not (Kim et al. 2008). Chou (2009) emphasizes that the entire feasibility of a road construction project is based on the accuracy of the cost estimation process. Accurate estimation of cost at the initial phase will aid construction managers in making necessary judgments on complicated construction projects (Petroutsatou et. al. 2012). Mahamid

(2013) also highlights the importance of conceptual cost estimation for clients, which is aiding in making appropriate budget decisions.

Even though it is extremely important to estimate cost accurately at the initial phase of construction projects, it is challenging because of the limited amount of information present (Soikov 2009). Developing countries also face additional cost estimation challenges because of their financial and systematic failures (Soikov 2009). Wilmot and Cheng (2003) stated that current estimation of highway construction cost does not consider the changes that might come as the project proceeds ignoring factors that affect future cost. Highway project managers should be extra cautious when making cost estimations because it is a key element for the success of a construction project (Wilmot and Cheng 2003). Although accurate estimation of construction costs is important to make necessary decisions at the initial phase of the project, several constraints make it challenging (Kim et al. 2008). These constraints include the inaccuracy of the scope defined at the beginning phase of a project, the limited amount of time given to the initial preparation of estimation of cost, and the struggle to find reliable data associated with incomplete final settled plan on the project at the early stage (Kim et al. 2008). Gardner et al. (2017) emphasized what makes the early estimation process a bit challenging is the fact that it is done before design and full project scope completion. To have better confidence in the initial estimate, the project scope should be clearly defined, and a reliable source of historical data should be available at the onset of the estimation process (Chou and O'Connor 2007). Both under and overestimation equally affect a project as inaccurate estimation affects the budget allocated and gives road project managers a bad reputation (Chou 2009). The limited amount of time given, inadequate data about the projects, and inexperienced estimators are some of the reasons behind inaccurate cost estimation (Chou 2009). The estimation process is plagued with political and environmental obstacles, less effective databases, and less effective ways used for the estimation (Mahamid 2013). The first estimation is always uncertain, but as the project goes by, additional information will be added which will increase the accuracy of the estimation (Le et al. 2019). Earthwork is expensive and the reliability of current estimation methods is unsatisfactory (Zitta et al. 2019). Few studies exist that specifically associate the cost with the earthwork (Zitta et al. 2019).

Various researchers have tried to implement different methods to solve the problem of inaccurate conceptual cost estimation of earthwork projects. The study done by Zitta et al. (2019), used a computerized system to perform the estimation of the cost of earthwork. Sthapit and Mori (1994) developed equations to forecast the cost of highway earthwork. Kim et al. (2004) prepared mathematical equations for each item of work to predict the intersection costs, which incorporates earthwork as the main part. Le et al. (2019), conducted a study that used historical bid data from the Iowa Department of Transportation. Asphalt binder PG 58-28 and PG 64-22 were selected as the study's bid items owing to their high proportions. Geographic information system (GIS) was used in the study. The data points in the study were identified by geographic coordinate system, latitude, and longitude in GIS according to their geographic locations. A geospatial interpolation feature in GIS was implemented to conduct the study, in which estimates of values at unknown locations were calculated by using values at known locations. The outcome of the study was a unit cost prediction visual map.

Wilmot and Cheng (2003) conducted a study with the main aim of preparing a model for the prediction of highway construction costs. The Louisiana Department of Transportation was used as the source of data for the study. The study used five sub-models for the prediction:

excavation & embankment, concrete pavement, asphalt pavement, reinforcing steel concrete, and structural.

Chou and O'Connor (2007) conducted a study to estimate highway construction cost at the early phase. Texas Department of Transportation provided the data. Using work break down structure, the work item was divided into simpler parts, which were then grouped into an 80-20 classification (with 80 being the frequently occurring items and 20 being the less appearing ones). An item-based model was employed, which took advantage of the web by posting information on a web-based system. The system followed three-tier architecture: client, database server, and web server tiers. The total estimation was found by summing the standard, non-standard, and contingency work items. Kim et al. (2008) stated that the highway operation handbook is widely used for the estimation of highway construction project costs. The problem with this method is it does not take into consideration important factors like location, which affect the project cost by causing variation from one project to the other (Kim et al. 2008).

2.3. Research Questions

This study was done as a systematic literature review, searching for articles involving conceptual cost estimates of highway earthworks to answer the specific research questions presented. Systematic literature review as a systematic way of deeply searching for pieces of literature that will answer originally defined questions using a formal and transparent searching method (Briner and Denyer 2012).

This study uses a systematic approach to find answers to the following two questions:

RQ 1: What are the methods which have been used in the conceptual cost estimation of highway earthwork construction and their limitations?

RQ 2: What are the factors affecting highway earthwork costs?

These questions were developed after careful review of several articles and after screening and changing several questions. The answer to these questions will be used as a base for our research.

2.4. Search Process

To find the most relevant articles several search words were utilized. Because there is limited research conducted on earthwork cost estimation, there was a need to modify the keywords, interchange word placement, and combine with other terms to find the right articles. A few of the keywords used were: “earthwork construction + cost”, “highway earthwork construction + cost”, “highway earthwork + cost”, earthwork + cost estimation”, highway + cost estimation”, “construction/highway/earthwork + cost”, and “highway + preliminary cost estimation.”

Various research databases were used to conduct the research. (Falagas et al. 2008), compared different medical databases and stated that this era of technology gave rise to the World Wide Web (WWW), which created an important ground for the growth of database systems. They emphasized that these databases made the mining of data on certain topics simpler. The databases also made deep evaluations and assessments of citations possible (Falagas et al. 2008). The databases used to extract information to answer the research questions for our study were: American society of civil engineers, google scholar, science direct, Scopus, and other miscellaneous databases.

Inclusion and exclusion criteria were applied while searching for the relevant articles: the included articles should specifically include cost estimation, the scope should be limited to the highway work and specifically to the earthwork, search should be specified to research journals, conference papers, or peer-reviewed articles, and the articles should show the method used for estimating highway, specifically earthwork cost and specify factors affecting the cost. Excluded articles include those that did not include the term highway, articles whose focus is not estimating, and articles that neither gave the methods of estimating nor the factors affecting cost.

The articles identified passed several stages of screening. After choosing the appropriate databases, the first stage was screening the articles based on the title. After collecting articles based on their titles, duplicate or unnecessary papers were removed by careful screening. The next step was screening by abstract. All abstracts of the selected articles were read. Based on their abstract, there were a significant number of papers that were excluded, since they no longer satisfied the need of the research questions. The final screening was done through reading of the papers entirely. The papers left after this last screening were used to answer the research questions of the systematic literature review.

2.5. Cost Estimation Methods Used for the Conceptual Stage of Highway Projects

The critical analysis of the systematic literature review gave answers to the two research questions presented.

RQ1: What kinds of methods are used for the conceptual cost estimation of highway earthwork construction and what are their limitations?

There were only limited number of articles conducted on the estimation highway earthwork construction. Eight papers are listed here to answer the research question one. Zitta et al. (2019), developed a computerized system to estimate the cost of highway earthwork. The study took place on one campus in Niger. To perform the study, a mass haul diagram was used as an input. The main aim of the study focused on the usage of computerized systems to perform the estimation of the cost of earthwork. A program code was developed to calculate volume. The volume at a specific fixed location on the mass haul diagram was then multiplied by the unit cost to estimate the total cost value. This computerized system gave rise to an estimation graph. An estimation graph was also done manually and compared with the one that was done by a computerized system. Although the two gave similar results, it is stated that the one done manually is labor-intensive and time-consuming, whereas the computerized one is less time-consuming and labor-intensive. Even though the computerized method is ideal, it is recommended to adjust the codes to give a better estimate and to collect the result in a well-organized database to make it easy to access and receive a public review.

Sthaphit and Mori (1994) developed a mathematical model to estimate highway earthwork costs. The study was conducted on Nepal highway earthworks. The main aim of the study was to forecast the cost of highway earthworks using developed equations. The estimated volume was validated, and the prepared model was tested statistically. Although the model proved to be effective, the study stated some limitations. Only a 100m-long road segment was used for the hill factor, so the study recommended to use other length road segments to improve the model.

Kim et al. (2004) also performed a research study with the main aim of estimating highway intersection costs. These intersection items in the study included earthwork cost, pavement cost, and right-of-way cost. Using these items, only the fill earthwork volume was predicted. Then, Kf, which represents the fill cost per cubic meter, was multiplied by the volume to estimate the total earthwork cost. However, when performing the prediction for the earthwork cost, the only component used was the fill component, so the researchers recommend in the future to also use cut components to make the study complete.

Asmar et al. (2011), presented a statistical analysis method which resembles the program evaluation review technique (PERT), that incorporates work breakdown structure to estimate the cost of roadway projects. The goal of the study was to estimate the cost at the early phase when little information is known about the project using statistical methods. The research used groups of roadway projects for the analysis: major, allowance, and structure units. These all were added to give the base cost estimate. Design and other contingencies were considered. A simplified work breakdown structure (WBS) was used in the study and the entire analysis showed resemblance with PERT analysis. A cost estimating template was utilized for easy insertion of data and the final estimate was cross validated, which gave promising results. However, one of its limitations is it gave less focus for contingency items and road incidentals, so the study recommended better collection of data in future studies.

Mahamid (2013) used a regression model to predict the future cost of road projects, in which earthwork cost was calculated being one of the main components. Saudi Arabian road projects were used as the source of data. Using quantity and size, five different models were developed to do the prediction. Three quantity variables and two size variables were used to build the model. The quantity variables used were earthworks, base works, and asphalt works; the size variables used were, road width and length. The first model in developing the equation used all the three quantity variables together. The second model created models separately for the three items. Asphalt work was chosen in the end because of the good correlation it showed. The third model used the size variables width and length together. The fourth model chose and took separately the size variable length as an independent variable owing to its significance. The fifth and last model used the interaction model of the width and length of the size. The study showed promising estimation results. However, the study recommended more work on improving the quality of the database and frequent checks and modifications of the unit prices and the model itself. The study also advocated the use of more variety of independent variables.

Kim et al. (2008) proposed a two-phased model for the estimation of road cost in which earthwork was considered as the major part. The model is called two-phased because it has two stages: the first stage is the initial planning phase and the second is the design phase where the drawings were prepared, and calculation of quantities was conducted. Different estimation models were applied for the two stages. For the first, a model called the case-based reasoning model, which bases its prediction on past similar models, was used. There is a limited amount of design information present at this initial phase, so past similar models were used. The second method took advantage of the information gathered after the design. It presented an estimation model which is based on quantity. After data was collected from a variety of projects, earthwork, drainage, and pavement were found to be significant to build the model. In this study, another quantity estimation model known as classification and regression trees (CART) was used specifically for earthwork and drainage since it was not able to read quantity from the drawings for those unit items. Although the study was helpful, it recommended the addition of bridge and tunnel on the estimation model to better predict the cost of road works.

An early estimation of roadway project cost using simple ways like a calculator and a simple computer program was suggested by Mahamid and Bruland (2010). The study used West Bank projects of 2005-2008 as the sources of data (in which the inflation index for the year 2008 was utilized to bring the data to that year). Multiple linear regression models were developed for the three work components: earthworks, base course works, and asphalt works. Historical databases and different statistical methods were used for the extraction and analysis of data. The units of the three dependent variables used were cost, cost per meter, and cost per meter square. They were presented as a function of project properties like road width, length, and soil condition. For asphalt and base course, the regression models used all three (cost, cost per meter length, cost per meter square), whereas the earthwork model used only the first two (cost and cost per meter length). The third cost per meter square was excluded because of its less significance in the earthwork model. Several models were prepared in the end under each group. The result indicated that the linear regression model of the road projects which used earthwork as the dependent variable was significant. This approach is helpful in accurately estimating early construction costs, but it is stated that its accuracy is still dependent on the preparation and elaboration of the details of the estimation process.

Sodikov (2005), used the method called artificial neural network (ANN) method in cost estimating of road projects specifically in developing countries. Rock database, which is populated by developing countries' road data, was used as a source of data. Poland and Thailand, having more data sets, were chosen for the study. Since the rock database was unable to provide earthwork and site preparation, the RCC model was created for those quantities. Using the model, the volume was calculated as a function of topography type and road width. Then, to calculate the cost, the regression model was used. Two ANN models were utilized to do the estimation. The study successfully used the ANN method to estimate the cost of future road projects in developing countries. It was recommended in the study to add more variety of datasets in addition to the earthwork and overlays. Finding more ways of enhancing the ANN model was also recommended.

2.6. Factors Affecting Highway Earthwork Construction Cost

RQ2: What are the factors affecting earthwork costs?

In their study for cost estimation of highway earthwork, Staphit and Mori (1994), identified factors influencing the cost as hill factors, slope factors, and soil type factors. According to Le et al. (2019), location effect and the time change were identified as the factors affecting the cost. From the two, the location factor was noticed to affect the cost more in this study in which the study specifically stated how location has a high impact on the accuracy of unit price prediction. Wilmot and Cheng (2003), in their study of cost estimation model preparation, identified the following three group of factors that influence the price: input contributing variables (manual labor, raw material, and equipment), contract feature variables (quantity and contract period), and bid variable (bid size). The study showed that the contract feature variable, specifically the quantity, has the highest influence on the price. The study also assessed the location factor and made a statement that the place of construction has an enormous impact on the price. According to Mahamid (2013), the quantity and size, specifically road width and length of the project, were seen to highly affect the prediction of future cost of roadworks. According to the study by Hegazy et al. (1998). St. John’s department of agriculture assisted in the provision of soil data dependent on location for the study. A couple of factors that affect the cost of highway projects were listed in the study, as the magnitude of the project, time, and place of the construction. In building the model, these factors were used as independent variables, whereas the cost was used as the dependent variable (Hegazy et al. 1998).

Kim et al. (2008) in their preparation of a two-phased model to estimate highway cost, point out the influencing factors are location, contract type, and road width. Mahamid and Bruland (2010), developed a regression model for conceptual cost estimation of road work projects. In this study properties like road width, length, haul distance, terrain condition, and soil condition were seen to greatly influence the cost. Additionally, Sodikov (2005) conducted a study that estimated future road costs in developing countries. In the study, the factors which have a high influence on the cost of earthwork and overlays were identified as earthwork volume and pavement width.

2.7. Discussion of Results

In response to research question 1, “What kinds of methods are used for the conceptual cost estimation of highway earthwork construction and what are their limitations?”, the results gathered from the collected articles are discussed. There were many articles found regarding cost estimation of road projects, but only a few articles focused specifically on earthwork. Few papers used a computerized system to estimate the cost of highway earthwork. Others used mathematical models to estimate highway earthwork costs, using developed mathematical equations. Other studies multiplied fill factors with the volume to predict of the cost of highway earthwork. Statistical methods, which closely resemble program evaluation review techniques, were also applied to some studies to estimate the cost. Different kinds of models were implemented to estimate costs, including regression model, multiple regression model, and case-based reasoning model. An artificial neural network (ANN) was also implemented to estimate road costs in developing countries. These methods showed promising results in achieving the goal of estimating the cost at the conceptual stage. Even though the methods highly assist the conceptual cost estimation of highway earthwork, scarcity of data is seen in most studies and more data is recommended to better enhance the results. It is suggested to improve database quality and implement more frequent inspection of unit prices. Even for the studies which draw models, frequent inspection is recommended, and additional data is suggested to enhance results. More preparation and elaboration of details of the estimation process are vital. The computerized systems need more adjustment of codes, and the studies recommend collection/preparation of a well-organized database to better present the results. Only a few articles have been written on methods of estimating earthwork cost and there is presently a limited amount of data in the initial stage of construction. For this reason, it is recommended to find an appropriate cost estimation method for highway earthwork construction at the conceptual stage.

In response to the second question, “What are the factors affecting earthwork cost?”, different researchers suggested different factors that might affect earthwork cost: Earthwork volume, pavement road width, road length, haul distance, terrain condition, soil condition, location, contract type, quantity, time change, soil type, hill, and slope factors were the factors identified from the collected studies. In most studies, the location factor was stated to have a high impact on the price of earthwork. Therefore, our study focuses on finding a method that considers location effect on cost of highway earthwork.

2.8. Conclusion

This paper conducted a systematic literature review of articles in which it reviewed studies done in the past to answer the research questions presented. The systematic literature review searched articles regarding the methods used for estimating the cost of highway earthwork and it assessed the factors that affect the price. It discussed the variety of methods used to predict the costs and it pointed out the limitations associated with each study. According to the research, it was found that the amount of literature done specifically on highway earthwork cost estimation is limited and even the ones which were found have a variety of limitations. So, there is a need for a new method that can help in accurately estimating the cost of highway earthwork construction. And the result of the second question showed a variety of factors that affect the cost. Location factor was repeatedly stated as a major influencing factor. In conclusion, this research shows that there is a need for a new method that will aid the accurate estimation of highway earthwork costs, filling the gap presented in the above studies. Future studies should also consider the effect of location because it is important to the final cost estimation.

2.9. References

Asmar, M. E., Hanna, A. S., & Whited, G. C. (2011). New approach to developing conceptual cost estimates for highway projects. Journal of construction engineering and management, 137(11), 942-949.

Bogenberger, C., Dell’ Amico, M., Fuellerer, G., Hoefinger, G., Iori, M., Novellani, S., & Panicucci, B. (2015). Two-phase earthwork optimization model for highway construction. Journal of construction engineering and management, 141(6), 05015003.

Briner, R. B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. Handbook of evidence-based management: Companies, classrooms and research, 112-129.

Chou, J. S. (2009). Generalized linear model-based expert system for estimating the cost of transportation projects. Expert Systems with Applications, 36(3), 4253-4267.

Chou, J. S., & O'Connor, J. T. (2007). Internet-based preliminary highway construction cost estimating database. Automation in Construction, 17(1), 65-74.

Easa, S. M. (1988). Selection of roadway grades that minimize earthwork cost using linear programming. Transportation Research Part A: General, 22(2), 121-136.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. The FASEB journal, 22(2), 338-342.

Gardner, B. J., Gransberg, D. D., & Rueda, J. A. (2017). Stochastic conceptual cost estimating of highway projects to communicate uncertainty using bootstrap sampling. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 3(3), 05016002.

Hegazy, T., & Ayed, A. (1998). Neural network model for parametric cost estimation of highway projects. Journal of construction engineering and management, 124(3), 210-218.

Hola, B., & Schabowicz, K. (2010). Estimation of earthworks execution time cost by means of artificial neural networks. Automation in Construction, 19(5), 570-579.

Jayawardane, A. K. W., & Harris, F. C. (1990). Further development of integer programming in earthwork optimization. Journal of construction engineering and management, 116(1), 18-34.

Kim, D. Y., Kim, B., & Han, S. H. (2008, June). Two-staged early cost estimation for highway construction projects. In ISARK, The 25th International Symposium on Automation and Robotics in Construction. Institute of Internet and Intelligent Technologies.

Kim, E., Jha, M. K., & Schonfeld, P. (2004). Intersection construction cost functions for alignment optimization. Journal of transportation engineering, 130(2), 194-203.

Kim, H., Bang, S., Jeong, H., Ham, Y., & Kim, H. (2018). Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. Automation in Construction, 92, 188-198.

Le, C., Le, T., Jeong, H. D., & Lee, E. B. (2019). Geographic information system–based framework for estimating and visualizing unit prices of highway work items. Journal of Construction Engineering and Management, 145(8), 04019044.

Mahamid, I. (2013). Conceptual cost estimate of road construction projects in Saudi Arabia. Jordan Journal of Civil Engineering, 7(3), 285-294.

Mahamid, I., & Bruland, A. (2010, April). Preliminary cost estimating models for road construction activities. In Proceedings of the FIG Congress.

Parente, M., Correia, A. G., & Cortez, P. (2014). Use of DM techniques in earthworks management: a case study. In Earthwork Project Management, Slope Stability Analysis, and Wave-Based Testing Techniques (pp. 1-8).

Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., & Pantouvakis, J. P. (2012). Early cost estimating of road tunnel construction using neural networks. Journal of construction engineering and management, 138(6), 679-687.

Sodikov, J. (2005). Cost estimation of highway projects in developing countries: artificial neural network approach. Journal of the Eastern Asia Society for Transportation Studies, 6, 1036-1047.

Sodikov, J. (2009). Road cost models for prefeasibility studies in developing countries. Journal of infrastructure systems, 15(4), 278-289.

Sthapit, N., & Mori, H. (1994). Model to estimate highway earthwork cost in Nepal. Journal of transportation engineering, 120(3), 498-504.

Wilmot, C. G., & Cheng, G. (2003). Estimating future highway construction costs. Journal of Construction Engineering and Management, 129(3), 272-279.

Zhang, Y., Minchin Jr, R. E., & Agdas, D. (2017). Forecasting completed cost of highway construction projects using LASSO regularized regression. Journal of Construction Engineering and Management, 143(10), 04017071.

Zitta, N., Aliyu, M. R., Peter, E. U., & Ayuba, Y. (2019). Estimating Cost of Earthwork Using Computerized Visual Display Approach. ATBU Journal of Science, Technology and Education, 7(1), 138-148

3. EXPLORATORY DATA ANALYSIS OF HIGHWAY EARTHWORK BID ITEMS

3.1. Introduction

Major activities in earthwork construction include the following: excavating, transporting, filling, compacting, reusing, temporary distribution of items, dumping soil materials, and pits material procurement (Bogenberger et al. 2015). The digging of earthwork involves excavating and transporting up to tens of millions of cubic meters of materials using different types of equipment, whereas filling involves transporting and filling various earth materials into already prepared cut areas or piling them for bridges or manmade passageways (Bogenberger et al. 2015). In the digging process, earth materials are classified based on their geological origin and are grouped to be prepared for their use (Bogenberger et al. 2015).

A mass haul diagram is one of the most widely used diagram in an earthwork, with the Y-axis as the cumulatively added volumes and the X-axis as the list of the stations (Zitta et al. 2019). The total volume is the summation of the cut which has a positive value and the fill which has a negative value along the station line (Zitta et al. 2019). According to Bogenberger et al. (2015), mines, burrow pits, and sites where the earth materials are dumped have certain limitations in which they cannot carry more earth material quantities than they could handle. Thus, their capacity should be carefully studied beforehand according to the study.

A high amount of cost is invested in an earthwork during the process of digging, filling, and hauling on a certain path (Zitta et al. 2019). The main challenge is finding the minimal cost associated with the earthwork activities (Bogenberger et al. 2015).

Conceptual cost estimation is the kind of estimate that is prepared where the specific details of a project are not known, and the design is not finalized (Asmar et al. 2011). A detailed estimate is done at 100% design completion, whereas a conceptual cost estimate is done when the design is at the initial stage, or close to 30% completion (Asmar et al. 2011). The study emphasized it is necessary to use past project information as a base due to the lack of information on the project at this stage. To predict the cost at the early phase of a construction project, a higher intellectual capability is expected from the individual predicting and good-quality historical data must be present (Liu and Zhu 2007). The availability of a scarce design information at the beginning phase will force the quantity surveyor to base their estimation on personal assumptions, which can be misleading (Liu and Zhu 2007). It is important to estimate a construction bid accurately because if estimated inaccurately, it could lead to enormous crises in terms of finance and reputation of stakeholders (Biruk et al. 2017). This study also stressed that estimating is a big responsibility on the estimators' side not only in terms of keeping accuracy but also finishing on time (Biruk et al. 2017). Predicting the cost at the early phase of construction will highly benefit state highway agencies since they are responsible for early authorizing and making important rulings regarding the funding of a project (Asmar et al. 2011). The study emphasized that any stakeholder responsible for early budgeting of project cost can use this estimate as a starting point to aid the funding decision. Budgeting is a necessary aspect of a project implemented by the owner, which affects the scope of the project (Roxas et al. 2019).

The estimate at the conceptual stage of the project not only includes the actual construction costs, but also the non-construction costs which includes the permit, engineering, and incidental costs (Asmar et al. 2011). Accurate early estimation guarantees the accomplishment of a project and is beneficial to the different stakeholders in the construction industry; for clients, it aids in better control of funding, and for contractors, it increases revenue gained from each project (Roxas et al. 2019). The study emphasized that estimating cost early is usually prone to error since the estimation is based on rough theoretical engineering. Various unpredicted events may occur as the project proceeds, including sudden market price rises and unexpected interruptions (Roxas et al. 2019). The study stated that, when early cost estimation is calculated, these factors are not usually taken into consideration because they do not appear at the initial stage of the project. Before contractors decide on the participation of a bid, they take two things into consideration: first, the profit they will get in terms of money and reputation, and second, the loss they will acquire if their bid is not accepted (Biruk et al. 2017). Because of the risk, they usually limit the number of bids they prepare during a specific time period (Biruk et al. 2017). The study continued by stating that contractors should be cautious when deciding to bid on certain projects in which gut feeling or experience might not be enough. The prospective bidder should base their bid on a reliable way to avoid financial losses (Biruk et al. 2017). For this reason, there is a need for an effective way to assess the proper approach to be followed (Biruk et al. 2017).

Data of highway construction projects constitute several elements: the project description, total quantity measurement of the road quantity in terms of length, project expenditure, stream of traffic, topography, and environmental factors (Williams et al. 2009). State highway agencies continuously check and update these data to have better control of the projects (Williams et al. 2009). Data mining is a way in which the behavior of the primal data is studied and the specific trend the data is following is carefully analyzed to make it easily understandable (Parente et al. 2014).

Because of current age development in technology and better means of gathering information, large databases in construction are starting to flourish (Parente et al. 2014). These databases incorporate; the methods of construction, the amount of time and money invested in the construction process (Parente et al. 2014). Specifically, to earthwork, the amount of earth material that is excavated, hauled, and filled according to the design of the construction is included in the database (Parente et al. 2014). The databases are more developed since the current construction equipment is supported by technology which includes a uniform check of load-carrying capacity and equipment set up (Parente et al. 2014).

Empirical data analysis (EDA) is one of the most stable and important means by which data will be studied at the conceptual stage, where the different trends the data follows will be carefully analyzed and early hypotheses be made and checked (Behrens et al. 1997). After careful understanding of its properties, EDA plots the data graphically as a step of preparing it for analysis (Cox 2017). For any given data, the X-axis represents the exploratory data which is the independent variable, and the Y-axis represents the dependent variable (Cox 2017). Careful understanding of data throughout the cycle is an important input to the development of models (Behrens et al. 1997). Sufficient provision of data is necessary for EDA to make assumptions that will assist in the model-building process (Behrens et al. 1997). EDA has a clear approach in formulating distinct kinds of queries, assumptions, and conclusions for each step of the process (Behrens et al. 1997). The primary goal of EDA is knowing the data, the pattern it is following, and in the end, choosing the statistical approach to implement it (Cox 2017).

Careful cleaning and organizing of data are necessary in EDA, for instance in the study performed by (Shrestha et al. 2017), in which the Montana Department of transportation provided the bid data for the project, the data of the specific projects were distinguished and divided based on the time frame as present and past data. Some redundancy was seen in this study in the bid data of the separate projects because they have the same source DOT. The data was then cleaned by removing items that do not seem to assist the analysis, like lump sum items and the items whose quantity has no constant relation with the cost. The study by Cox (2017) identified 2 outliers in the data: data entry errors and statistical outliers. They eliminated the first groups and kept the latest ones for the further analysis process.

The objective of the exploratory data analysis is to extract highway earthwork bid item data from the Iowa Department of Transportation's official site with the goal of cleaning and preparing the data. This will make it ready for further analysis of the kriging methods which will be discussed in the coming chapters. This study involves analyzing the pattern of the data, the relationship of variables, the identification and removal of outliers, and the transformation of skewed graphs into normal curves. The study will elaborate the data exploration procedure followed and the outcome of the analysis will be discussed the in the end. The paper will conclude by stating the implications of the study and giving recommendations for further studies.

3.2. Data Collection and Preparation

This study identified the source of data and explored the different aspects of the entire set collected. The highway earthwork cost data was acquired from the Iowa Department of Transportation (DOT). Iowa DOT was chosen because of the large amount of data carried by the state department database which aids the analysis. Upon an email request, the department provided bid data of cost of highway earthwork projects from the year 2015 to 2020. The bid items provided by the department were: embankment in place, excavation-cl10- roadway + borrow, excavation-cl12-roadway+boulder/rock fragment, excavation boulder and contractor furnish select treatment. Since the amount of data that can be acquired for the other bid items was minimum, only three bid items were chosen for this study; embankment in place, excavation boulder, and excavation-cl10-roadway+borrow. The earthwork cost data was provided in an excel format with attributes: name of the bid item, unit, the letting date, the project number, quantity, price, amount, county, work type, project location, and specific longitude and latitude positions. The major type of works presented are hot mix asphalt mill and overlay, hot mix asphalt pavement widening and resurfacing, grading, reinforced concrete box culvert replacement, Portland cement concrete pavement grading and replacing, surface treatment for roads and more. Iowa State has a total of 99 counties, and data was collected from all 99 counties. Table 1 shows the original sample of data acquired from the state department for the embankment-in-place bid item for the year 2015 and the summary for the total number of bid items used for this study.

Description	units	Project Number	Quantity	Price	ext amt	County	Work Type	Longitude	Latitude
EMBANKMENT-IN-PLACE	CY	STP-U-8190(631)--70-09	545	12	6540	BREMER	HMA RESURFACING WITH MILLING	92.4792	42.7236
EMBANKMENT-IN-PLACE	CY	STPN-141-4(28)--2J-14	438.2	27.1	11875.22	CARROLL	HMA PAVEMENT WIDEN/HMA RESURFC	94.7544	41.8769
EMBANKMENT-IN-PLACE	CY	BRFN-038-2(33)--39-16	2819	16.15	45526.85	CEDAR	GRADING	91.1328	41.6869
EMBANKMENT-IN-PLACE	CY	BRFN-031-3(8)--39-18	2561	12	30732	CHEROKEE	RCB CULVERT RPLC - TWIN BOX	95.6603	42.6183
EMBANKMENT-IN-PLACE	CY	STP-S-C020(94)--5E-20	567	13.3	7541.1	CLARKE	BRIDGE REPLACEMENT - PPCB	93.9822	40.9475
EMBANKMENT-IN-PLACE	CY	FM-C022(079)--55-22	4520	5.2	23504	CLAYTON	PCC PAVEMENT - GRADE/REPLACE	92.8303	42.995
EMBANKMENT-IN-PLACE	CY	BRFN-061-2(89)--39-29	1982	15.9	31513.8	DES MOINES	RECONSTR - BRIDGE WIDENING	91.1567	40.7456
EMBANKMENT-IN-PLACE	CY	BRFN-017-3(34)--39-40	400	30	12000	HAMILTON	BRIDGE DECK OVERLAY	93.8158	42.4492
EMBANKMENT-IN-PLACE	CY	NHSN-030-1(144)--2R-43	13	100	1300	HARRISON	BRIDGE DECK OVERLAY	96.0953	41.5511
EMBANKMENT-IN-PLACE	CY	NHSN-030-1(144)--2R-43	39	45	1755	HARRISON	BRIDGE DECK OVERLAY	96.0953	41.5511
EMBANKMENT-IN-PLACE	CY	MP-003-2(708)137--76-46	40	28.5	1140	HUMBOLDT	RCB CULVERT - REPAIR	94.0678	42.7317
EMBANKMENT-IN-PLACE	CY	STPN-009-1(46)--2J-60	110	54.6	6006	LYON	MICROSURFACING	95.8803	43.4328
EMBANKMENT-IN-PLACE	CY	BROS-C071(76)--5F-71	444	10	4440	O BRIEN	RCB CULVERT RPLC - TWIN BOX	96.2583	43.0589
EMBANKMENT-IN-PLACE	CY	BRFN-163-1(87)--39-77	150	30	4500	POLK	BRIDGE DECK OVERLAY	93.5397	41.6006
EMBANKMENT-IN-PLACE	CY	BROS-1945(811)--8J-77	3386	19	64334	POLK	RCB CULVERT NEW - SINGLE BOX	94.4453	41.5294
EMBANKMENT-IN-PLACE	CY	STPN-010-1(79)--2J-84	90	54.6	4914	SIOUX	MICROSURFACING	96.1147	42.9972
EMBANKMENT-IN-PLACE	CY	STPN-069-5(105)--2J-85	1099	6.5	7143.5	STORY	HMA RESURFACING WITH MILLING	93.61	41.9794
EMBANKMENT-IN-PLACE	CY	NHSN-065-3(45)--2R-91	531	18	9558	WARREN	HMA PAVEMENT WIDEN/HMA RESURFC	93.5633	41.2856
EMBANKMENT-IN-PLACE	CY	BRF-001-4(43)--38-92	3419	8.4	28719.6	WASHINGTON	BRIDGE REPLACEMENT - PPCB	91.7161	41.2817
EMBANKMENT-IN-PLACE	CY	BROS-C099(77)--5F-99	752.4	17.5	13167	WRIGHT	BRIDGE REPLACEMENT - CCS	94.1678	42.8342
EMBANKMENT-IN-PLACE	CY	BRFN-071-4(47)--39-05	250	35	8750	AUDUBON	BRIDGE DECK OVERLAY	94.9378	41.755

	2015	2016	2017	2018	2019	2020	SUM
EMBANKMENT-IN-PLACE, CONTRACTOR FURNISH	174	195	178	162	167	172	1048
EXCAVATION, CL 10, RDWY+BORROW	216	234	219	218	192	234	1313
EXCAVATION, CL 12, BOULDER/ROCK FRAGMENT	62	53	58	46	48	45	312
						TOTAL	2673

Fig. 3.1: 2015 embankment in place bid item data & summary before outlier removal

A geographic information system (GIS) is the system used to explore the behavior of the data. Histograms and QQ plots on GIS were implemented to detect outliers and check the normality of the data set. Outliers in the data set must be removed because of the negative effect they have on the result of the analysis, but extra care must be taken not to remove the outliers in the region which are needed to give the complete result of the analysis (Cox 2017). Different statistical means were used to assess the behavior of the data. The presence of outliers and normality was checked by visual detection from the histogram and QQ plot and by paying careful attention to the values of descriptive statistical measures, mean, and the median. In addition to relying on GIS for outlier removal, Minitab software was used to clean any additional outliers present, in which the Grubbs test was applied. In that more outliers were detected and removed. Inflation was considered on all data points from 2015 to 2020 to bring the price values to 2021 the estimation period.

3.3. Outlier Detection and Removal

GIS is a great system that helps in assessing the behavior of data in relation to its spatial location. In the beginning, when placing the data points on GIS, seven bid items from the embankment and thirteen bid items from the excavation of roadway and borrow were found to be out of the border of the Iowa state. They were: two from embankment (2015), three from embankment (2016), two from embankment (2019), eight from excavation for roadway &borrow (2015), four from excavation for roadway &borrow (2016), one from excavation for roadway &borrow (2019). But since these points were out of the study area, they were detected and removed. After the data outside the border was screened out, outliers were detected in the bid items for each year and each bid item separately. Histograms and QQ plots were used to identify the values which are so far away from the rest of the data set. For instance, for the embankment bid item, the average value of the bid data was seen to be 15-20 but the outlier that was removed has a value of 200. Similarly, for excavation of roadway and borrow, the average value of the bid data for one specific year was 12-13 but the outlier that was removed has a value of 270. The outliers were removed to avoid leading to false conclusions. These extreme values were first identified using the histogram. QQ plot charts were then cross-checked on the original excel values before removing them to indicate that extra care was taken. After the removal of outliers using GIS chart histogram and QQ plot, there were still few outliers left that initiate the use of Minitab. The Grubbs outlier detection technique in Minitab was used to screen the remaining outliers. The final data used for the study after the outlier removal is presented in the figure below with the summary.

1	Letting Date	Project Number	Quantity	New Price	New amt	County	Work Type	Project Location	Longitude	Latitude
2	1/21/2015	STP-U-8190(631)--70-09	545	16.76	9133.71	BREMER	HMA RESURFACING WITH MILLING	IN THE CITY OF WAVERLY, 2ND AVE SW: FROM 4TH ST. SW TO 10TH	-92.4792	42.7236
3	1/21/2015	STPN-141-4(28)--2J-14	438.2	37.85	16584.83	CARROLL	HMA PAVEMENT WIDEN/HMA RESURFC	FROM 0.2 MILE EAST OF 4TH ST. IN DEDHAM, EAST TO	-94.7544	41.8769
4	1/21/2015	BRFN-038-2(33)--39-16	2819	22.55	63582.41	CEDAR	GRADING	APPROX 3 MILES N OF I-80 (6 LOCATIONS)	-91.1328	41.6869
5	1/21/2015	BRFN-031-3(8)--39-18	2561	16.76	42920.05	CHEROKEE	RCB CULVERT RPLC - TWIN BOX	FOUR MILE CREEK 1.1 MILES S OF CO RD L51	-95.6603	42.6183
6	1/21/2015	STP-S-C020(94)--5E-20	567	18.57	10531.84	CLARKE	BRIDGE REPLACEMENT - PPCB	APPROX. 3/4 MI. E. OF HOPEVILLE	-93.9822	40.9475
7	1/21/2015	FM-C022(079)--55-22	4520	7.26	32825.49	CLAYTON	PCC PAVEMENT - GRADE/REPLACE	PIKES PEAK RD: FROM GREAT RIVER RD TO STATE PARK	-92.8303	42.995
8	1/21/2015	BRFN-061-2(89)--39-29	1982	22.21	44011.91	DES MOINES	RECONSTR - BRIDGE WIDENING	US 61 OVER SPRING CREEK 3.7 MILES N. OF LEE CO. LINE	-91.1567	40.7456
9	1/21/2015	BRFN-017-3(34)--39-40	400	41.90	16759.10	HAMILTON	BRIDGE DECK OVERLAY	IA 17 AT E. JCT. US 20	-93.8158	42.4492
10	1/21/2015	NHSN-030-1(144)--2R-43	39	62.85	2451.02	HARRISON	BRIDGE DECK OVERLAY	OVER MISSOURI RIVER E. OF BLAIR NEBRASKA	-96.0953	41.5511
11	1/21/2015	MP-003-2(708)137--76-46	40	39.80	1592.11	HUMBOLDT	RCB CULVERT - REPAIR	8.1 MI EAST OF U.S. 169	-94.0678	42.7317
12	1/21/2015	STPN-009-1(46)--2J-60	110	76.25	8387.93	LYON	MICROSURFACING	CO. RD. L26 TO 0.9 MILE E. OF CO. RD. L40	-95.8803	43.4328
13	1/21/2015	BROS-C071(76)--5F-71	444	13.97	6200.87	O BRIEN	RCB CULVERT RPLC - TWIN BOX	L40: NW COR SEC 7 S 0.75 MI	-96.2583	43.0589
14	1/21/2015	BRFN-163-1(87)--39-77	150	41.90	6284.66	POLK	BRIDGE DECK OVERLAY	IA 163 OVER FOUR MILE CREEK	-93.5397	41.6006
15	1/21/2015	BROS-1945(811)--8J-77	3386	26.54	89848.32	POLK	RCB CULVERT NEW - SINGLE BOX	IN THE CITY OF DES MOINES: E. PAYTON AVE. FROM 120' WEST OF	-94.4453	41.5294
16	1/21/2015	STPN-010-1(79)--2J-84	90	76.25	6862.85	SIOUX	MICROSURFACING	FROM E. OF US 75 TO ALBANY AVE IN ORANGE CITY	-96.1147	42.9972
17	1/21/2015	STPN-069-5(105)--2J-85	1099	9.08	9976.55	STORY	HMA RESURFACING WITH MILLING	FROM JUST S. OF IA. 210 N. TO THE SCL OF AMES	-93.61	41.9794
18	1/21/2015	NHSN-065-3(45)--2R-91	531	25.14	13348.62	WARREN	HMA PAVEMENT WIDEN/HMA RESURFC	S JCT CO RD G58 TO US 69	-93.5633	41.2856
19	1/21/2015	BRF-001-4(43)--38-92	3419	11.73	40109.55	WASHINGTON	BRIDGE REPLACEMENT - PPCB	IA 1 OVER W. FORK CROOKED CREEK	-91.7161	41.2817
20	1/21/2015	BROS-C099(77)--5F-99	752.4	24.44	18388.92	WRIGHT	BRIDGE REPLACEMENT - CCS	ON 150TH STREET, OVER DRAINAGE DITCH, NW S32 T92N R25W	-94.1678	42.8342
21	2/17/2015	BRFN-071-4(47)--39-05	250	48.88	12220.18	AUDUBON	BRIDGE DECK OVERLAY	OVER BLUE GRASS CREEK 0.6 MILES N OF N JCT CO RD F32	-94.9378	41.755

1		2015	2016	2017	2018	2019	2020	SUM
2	EMBANKMENT-IN-PLACE, CONTRACTOR FURNISH	169	188	175	146	155	169	1002
3	EXCAVATION, CL 10, RDWY+BORROW	197	227	206	212	187	226	1255
4	EXCAVATION, CL 12, BOULDER/ROCK FRAGMENT	61	53	56	45	45	45	305
5							TOTAL	2562

Fig. 3.2: 2015 embankment in place bid items & summary after outlier removal

Outliers were detected by the Grubbs outlier detecting mechanism and studying the trend of histogram and QQ plot before taking the data for further analysis. Attached here is a sample figure for outlier removal of embankment 2019 data after the outlier was removed.

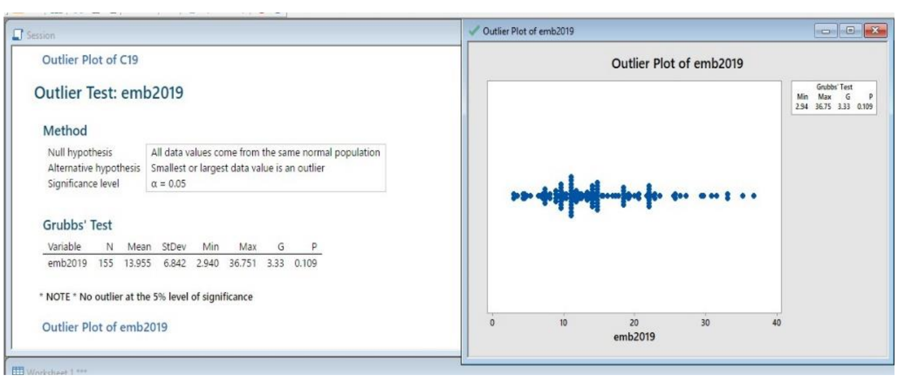


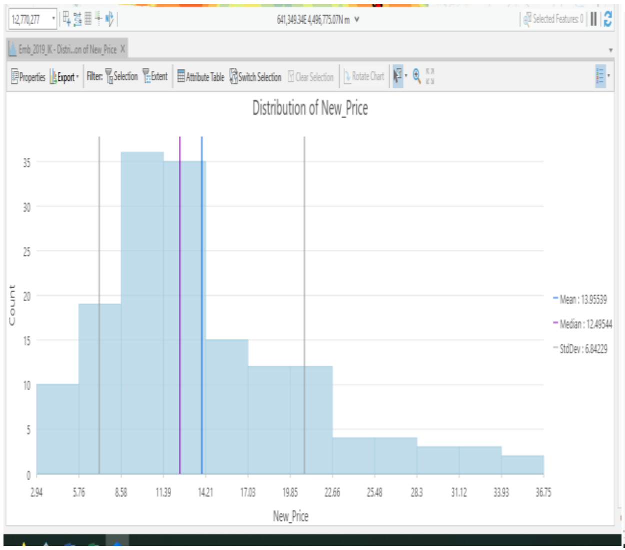
Fig. 3.3: Outlier detection using Grubbs

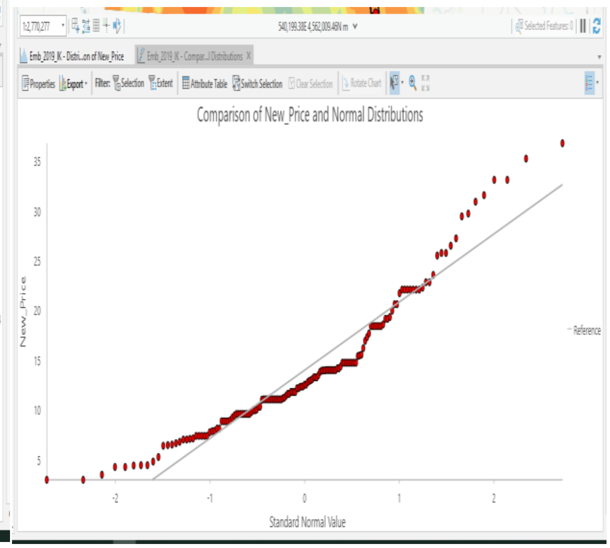
3.4. Normality Test and Transformation

The normality aspect of the data was similarly checked using histogram and QQ plot charts to assist the accurate prediction. The result from the analysis showed, almost all graphs were skewed to the right. Sample data is provided in Figure 1 for Embankment 2019 and Excavation Roadway and Borrow 2020. As shown in the figure, the histogram indicates that the data is skewed to the right. The right tail is an indication that only a few points have large prices. The visual perception of the histogram implies that the original data is not normally distributed. In addition to the visual judgments of the graphical representations of being bell curve, statistical descriptive measures, mean, median, and mode were also used to check normality. As it can be seen from figure 1, the mean and median do not have close values, thus supports the graphical assessment in which it is a clear indication for the distribution of data not being normal. And also, when noticing the QQ plot, since the plotted data points are not that close to the straight 45-degree line, it can be inferred that normality is not followed by the distribution of the original data points. So, from both the histogram and the QQ plot graphical representations and measure of central tendency, mean and median, it can be concluded that the data do not follow a normal distribution. This implies that there is a need for the transformation of the data into a normal distribution to give a better accurate prediction result.

Emb 2019

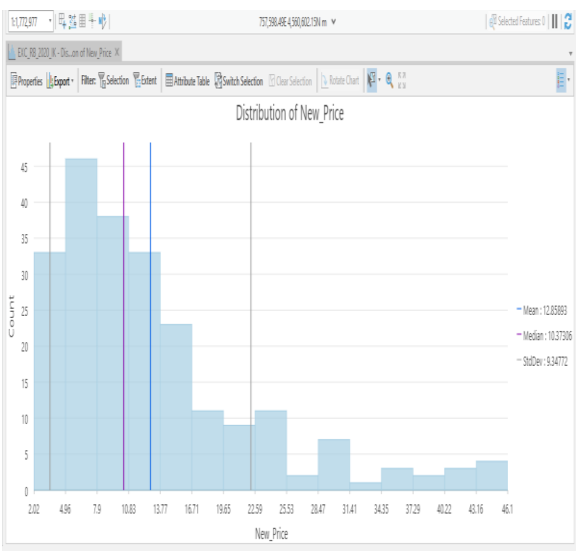
Histogram **QQ Plot**





EXC RB 2020

Histogram **QQ Plot**



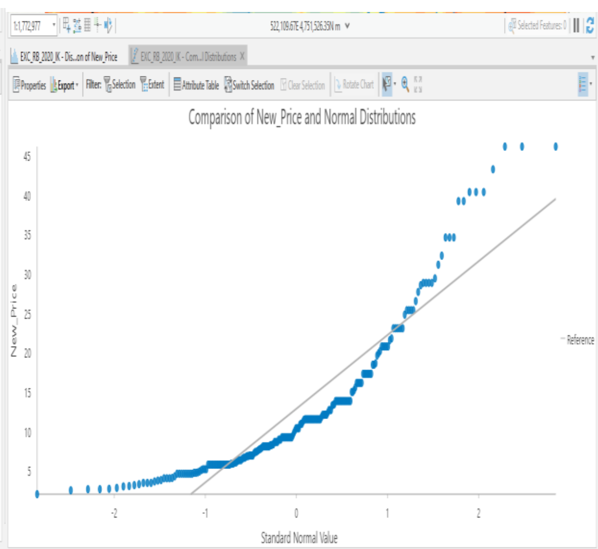


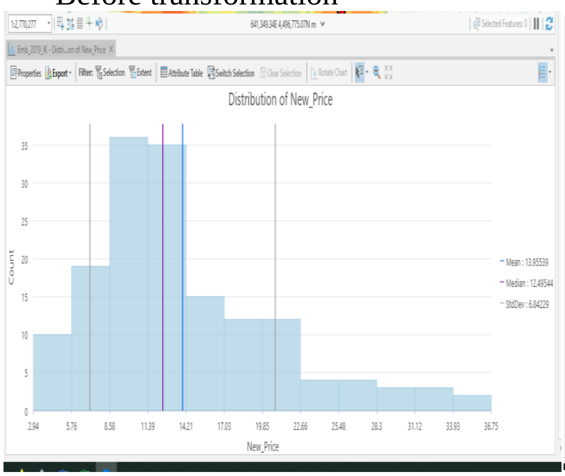
Fig. 3.4: Histogram and QQ plot of skewed data set before log transformation

Figure 2 shows the shape of two data sets from each unit item of work after log transformation was implemented. embankment 2019, embankment 2020, excavation boulder 2018, excavation boulder 2020, excavation roadway and borrow 2016, excavation roadway and borrow 2020 are the sample data sets shown in the figure. Outliers were removed and logarithmic transformation was applied on all the data sets for each year. The before and after logarithmic transformation is shown side by side for the histogram and QQ plot. As it can be inferred from the figure, after the transformation, the graphs are transformed into bell-shaped normal histogram curves and the mean and median values become closer.

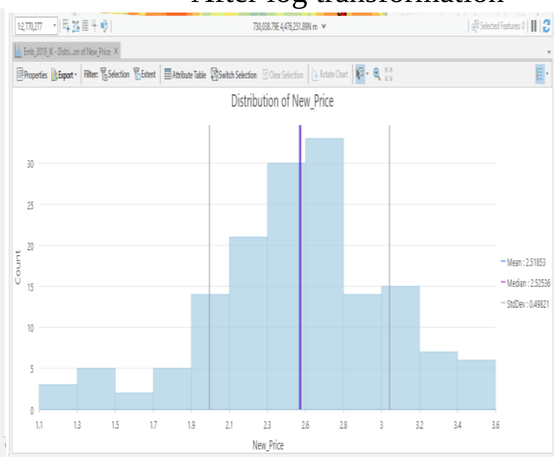
Embankment 2019

Histogram

Before transformation

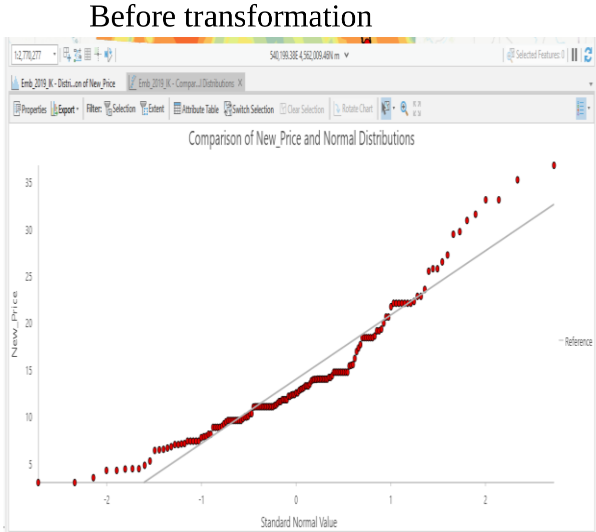


After log transformation



QQ Plot

Before transformation



After log transformation

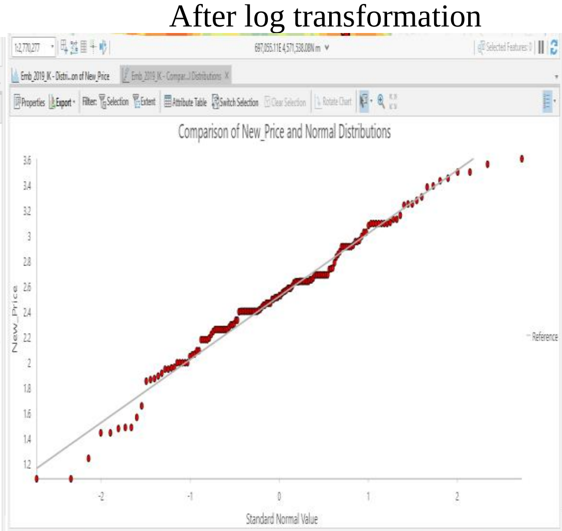
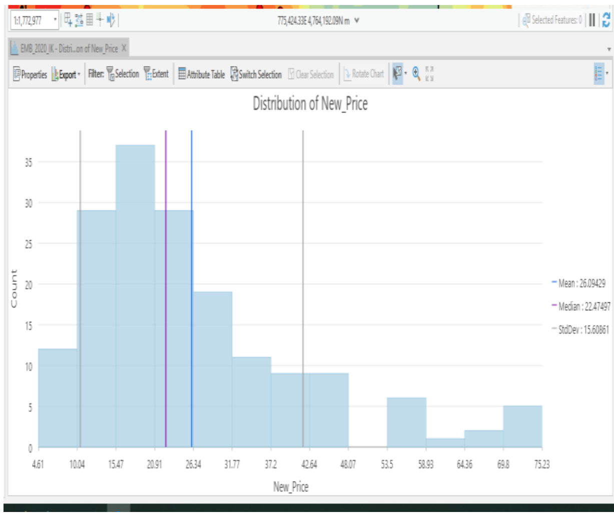


Fig. 3.5: Histogram and QQ plot of selected bid items after log transformation

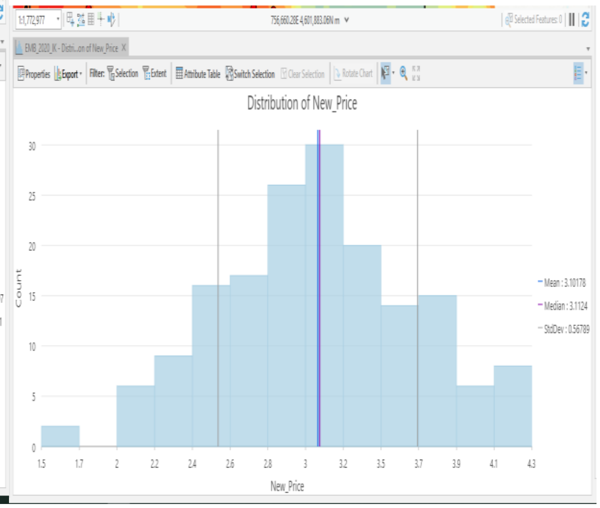
Embankment 2020

Histogram

Before transcformation

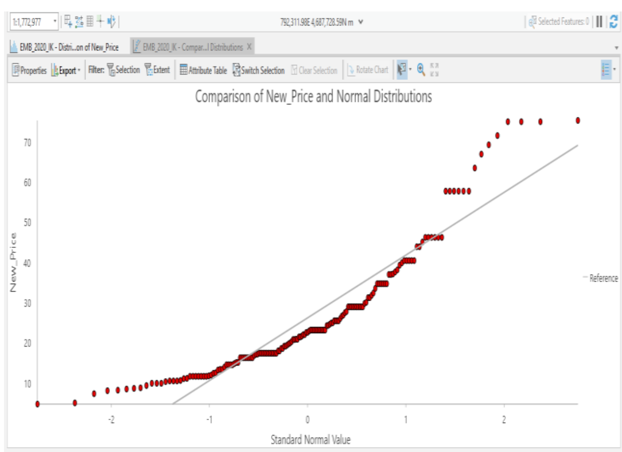


After log transformation



QQ Plot

Before transformation



After log transformation

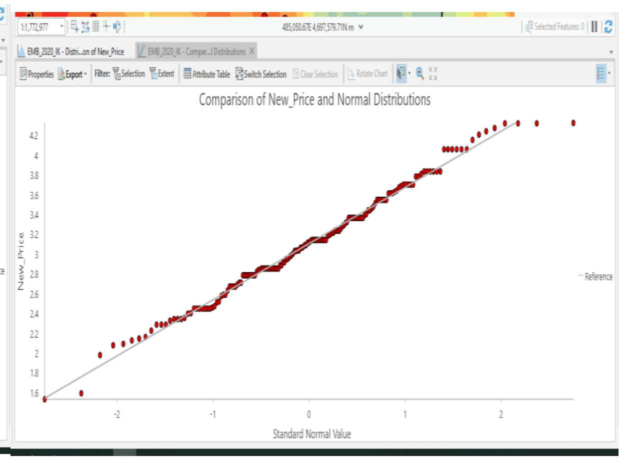
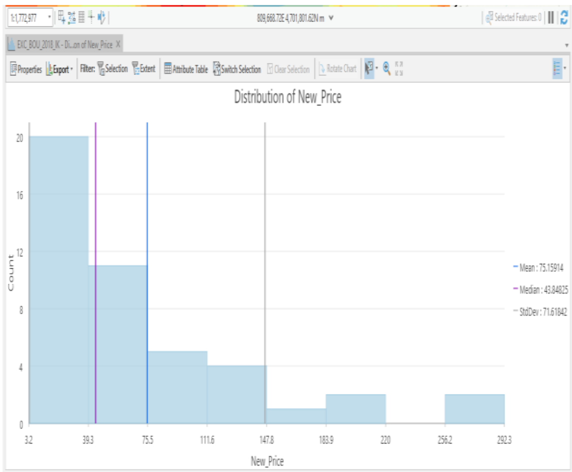


Fig. 3.5: Histogram and QQ plot of selected bid items after log transformation (continued)

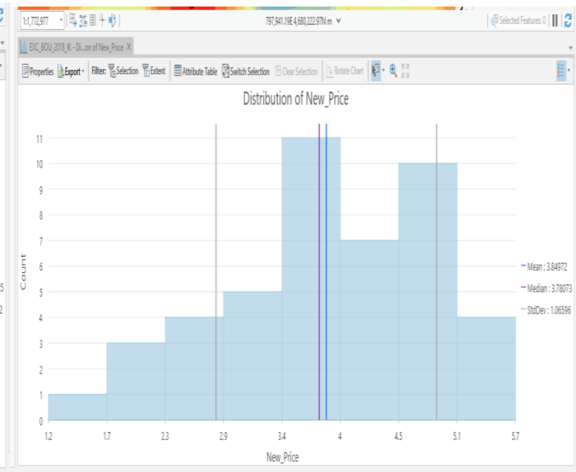
Excavation Boulder 2018

Histogram

Before transcribation

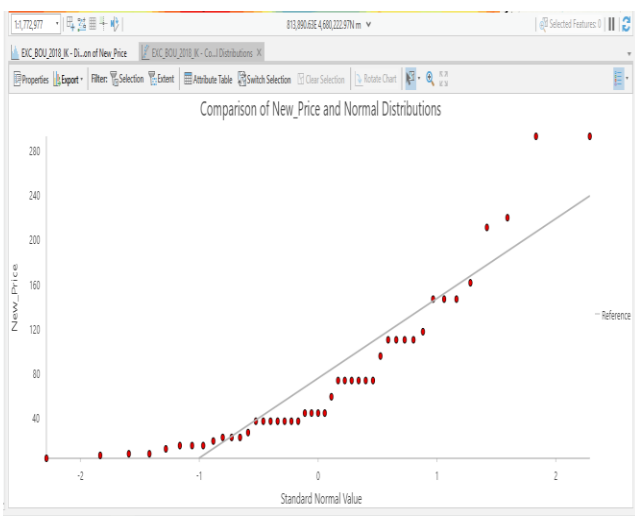


After log transformation



QQ Plot

Before transformation



After log transformation

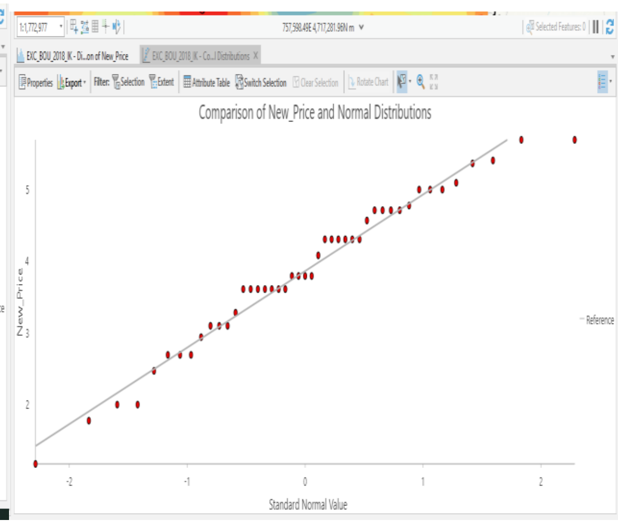
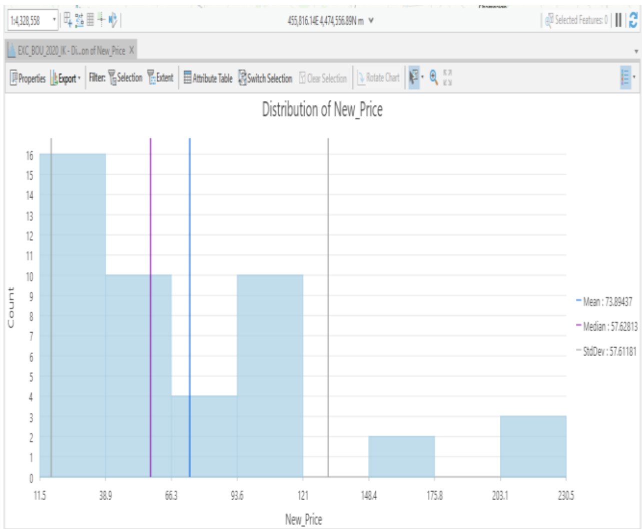


Fig. 3.5: Histogram and QQ plot of selected bid items after log transformation (continued)

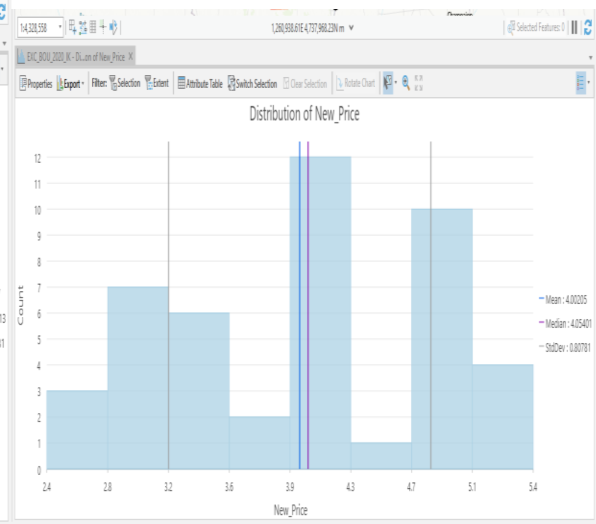
Excavation Boulder 2020

Histogram

Before transformation

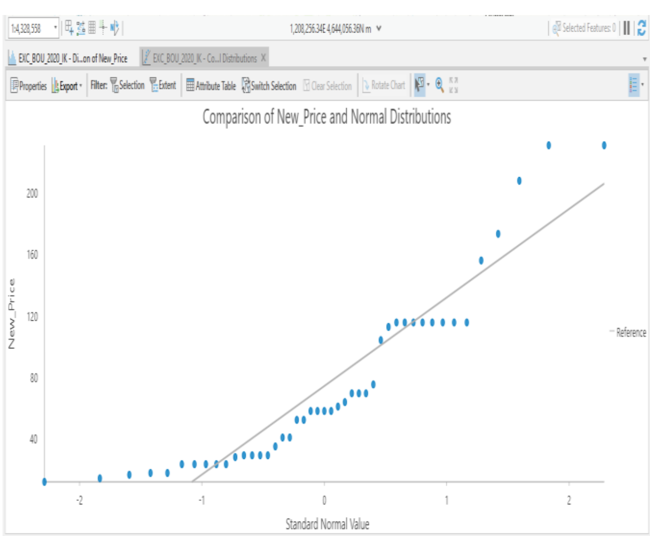


After log transformation



QQ Plot

Before transformation



After log transformation

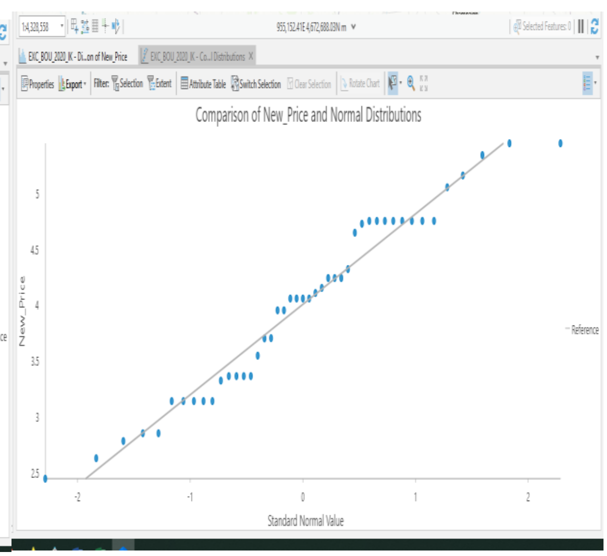
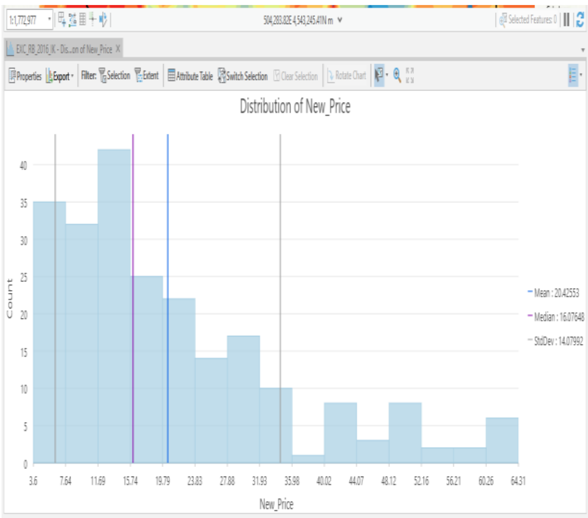


Fig. 3.5: Histogram and QQ plot of selected bid items after log transformation (continued)

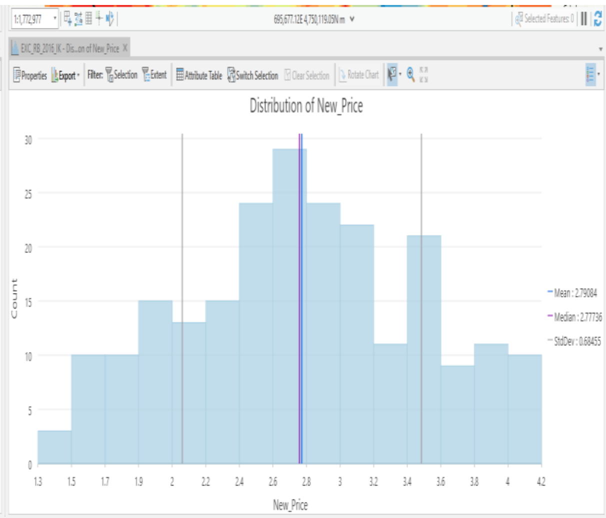
Excavation Roadway & Borrow 2016

Histogram

Before transformation

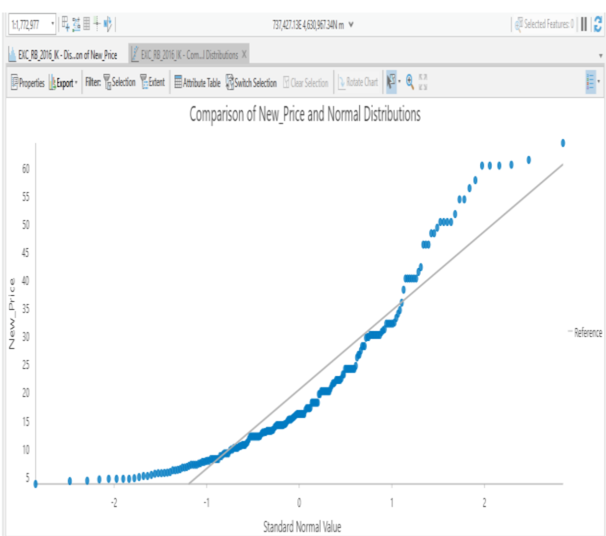


After log transformation



QQ Plot

Before transformation



After log transformation

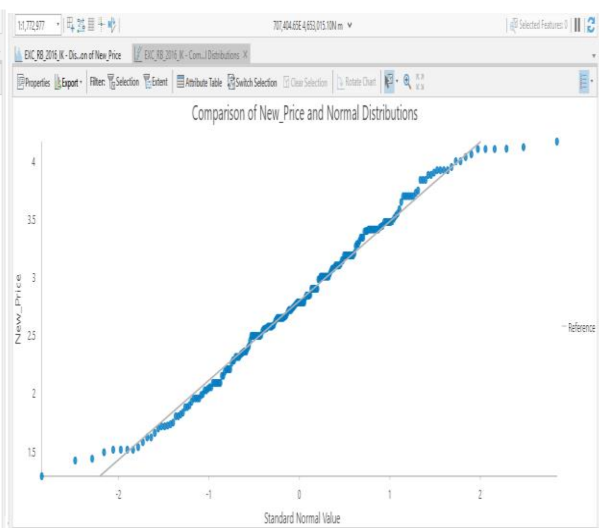
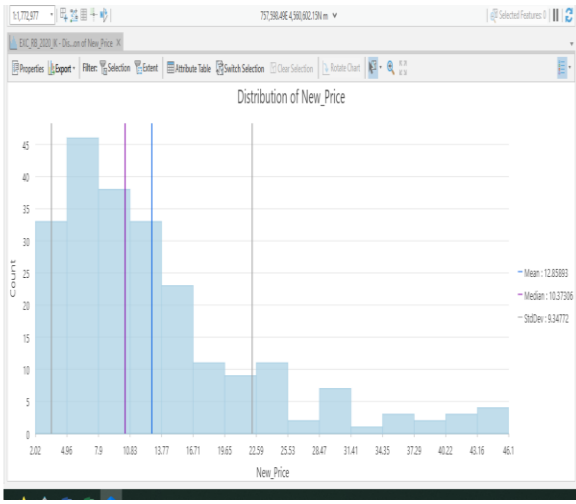


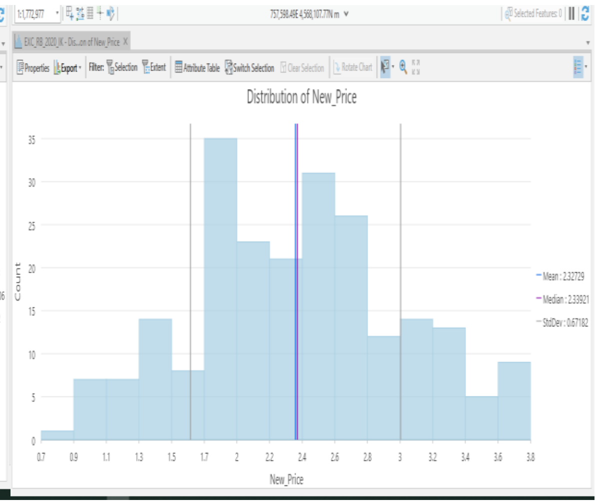
Fig. 3.5: Histogram and QQ plot of selected bid items after log transformation (continued)

Excavation Roadway & Borrow 2020

Histogram

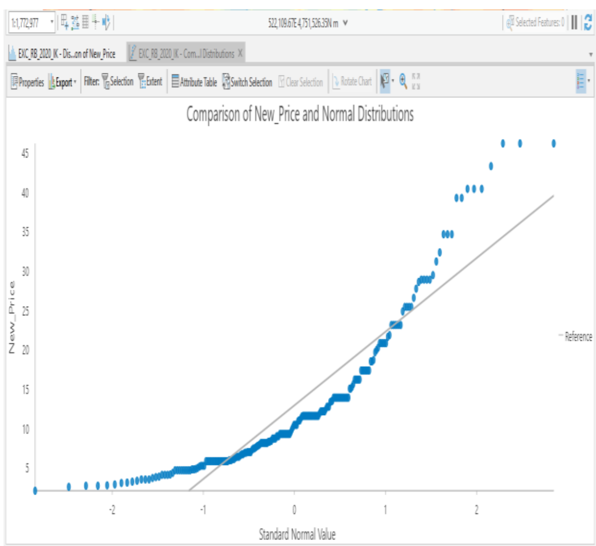
Before transformation   After log transformation





QQ Plot

Before transformation   After log transformation



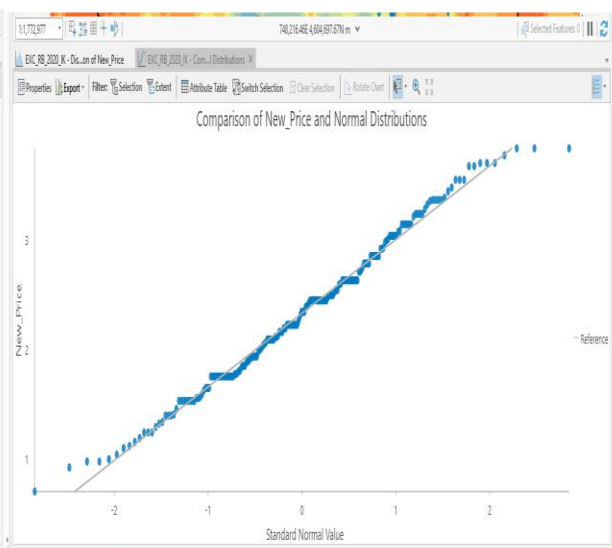


Fig. 3.5: Histogram and QQ plot of selected bid items after log transformation (continued)

After normal transformation, the data is ready for further analysis. This approach of cleaning data is important because, if not taken care of at earlier stages, it might mislead the interpolation results.

3.5. Conclusion

State highway departments have a high amount of historical data. It is necessary to study the trend of the past data to be able to predict the future. This section of the study involves studying the source and behavior of the data. Proceeding with analysis before cleaning the data leads to incorrect results. That is why exploratory data analysis is necessary. In this study highway earthwork cost bid data was collected from the Iowa department of transportation for the years 2015 to 2020. Geographic Information System (GIS) was used to study the behavior of the data. Outliers and data points outside the study border that deviate highly from the rest of the data were removed. The result found out that most data points were skewed to the right, which initiates the process of normalizing them. This study confirmed the necessity of studying the trend of the data before proceeding to the next stages. Although this analysis is helpful, it is limited only to three bid items. In the future, it is recommended to use more variety kinds of bid data to compare the results and increase the accuracy of the exploratory data analysis.

3.6. References

Asmar, M. E., Hanna, A. S., & Whited, G. C. (2011). New approach to developing conceptual cost estimates for highway projects. Journal of construction engineering and management, 137(11), 942-949.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. Psychological Methods, 2(2), 131.

Biruk, S., Jaśkowski, P., & Czarnigowska, A. (2017). Modeling contractor's bidding decisions. Procedia Engineering, 182, 91-98.

Bogenberger, C., Dell'Amico, M., Fuellerer, G., Hoefinger, G., Iori, M., Novellani, S., & Panicucci, B. (2015). Two-phase earthwork optimization model for highway construction. Journal of construction engineering and management, 141(6), 05015003.

Cox, V. (2017). Translating Statistics to Make Decisions. Translating Statistics to Make Decisions.

Liu, L., & Zhu, K. (2007). Improving cost estimates of construction projects using phased cost factors. Journal of construction engineering and management, 133(1), 91-95.

Parente, M., Correia, A. G., & Cortez, P. (2014). Use of DM techniques in earthworks management: a case study. In Earthwork Project Management, Slope Stability Analysis, and Wave-Based Testing Techniques (pp. 1-8).

Roxas, C. L. C., Roxas, N. R., Cristobal, J., Hao, S. E., Rabino, R. M., & Revalde, F. (2019). Modeling road construction project cost in the Philippines using the artificial neural network approach. In 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (pp. 1-5). IEEE.

Shrestha, K. J., Jeong, H. D., & Gransberg, D. D. (2017). Multidimensional highway construction cost indexes using dynamic item basket. Journal of Construction Engineering and Management, 143(8), 04017036.

Williams, R. C., Hildreth, J. C., & Vorster, M. C. (2009). Highway construction data collection and treatment in preparation for statistical regression analysis. Journal of Construction Engineering and Management, 135(12), 1299-1306.

Zitta, N., Aliyu, M. R., Peter, E. U., & Ayuba, Y. (2019). Estimating Cost of Earthwork Using Computerized Visual Display Approach. ATBU Journal of Science, Technology and Education, 7(1), 138-148.

4. CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK UNIT PRICE USING EMPIRICAL BAYESIAN KRIGING

4.1. Introduction

It is a challenging task to estimate the cost of construction bid items at the planning phase. Cost estimation is performed at various stages of a construction project, and it is an important aspect in assisting proper supervision (Hyari et al. 2016). Cost estimation appears at various stages of projects and the level of accuracy and care that must be taken varies with the kind of estimate (Hyari et al. 2016). There are various unknown factors and a minimum amount of information at the beginning phase of a construction project, which makes it hard for a conceptual cost estimate to be accurate (Elbeltagi et al. 2014). The later execution of the project is dictated by the decision made at the initial stage of the project (Elbeltagi et al. 2014). The successful completion of an infrastructure project is dependent on the conceptual cost estimate (Adel et al. 2016).

Conceptual cost estimates are prepared when the plan and definition of the project are not yet complete, and they are usually based on past historical data and models (Hyari et al. 2016). Unlike a conceptual cost estimate, a detailed cost estimate is done after more information is known about the project (Hyari et al. 2016). Gardner et al. (2016) also emphasized this point by defining conceptual cost estimate as the pioneer estimate prepared when there is insufficient detail and information about the design of a project. Further details of design and project definition will follow up in the consequent estimations made at later phases (Gardner et al. 2016).

Conceptual cost estimation has various benefits for the different stakeholders in a project (Adel et al. 2016). It assists the owner in better controlling and making necessary decisions concerning budget (Adel et al. 2016). For the designer, it aids in creating a design that aligns with the budget of the client (Adel et al. 2016). Being the first step, conceptual cost estimate not only assists the assessment of a project's practicability but also its successful completion (Cheng et al. 2009). The study further emphasizes that this process highly assists clients in knowing the practicability of the project in terms of cost ahead of time.

Cost estimation is a crucial construction project management activity that affects the project's aspects like design, budget, and construction (Cheng et al. 2009). The main purpose in estimating the cost of a project at the beginning phase is to decide on the budget and predict the cost of the completed project (Hyari et al. 2016). Highway construction projects are associated with different factors which affect the cost (Elbeltagi et al. 2014). These factors need to be accurately calibrated to better assist in the conceptual cost estimation of new projects (Elbeltagi et al. 2014). Gardner et al. (2016) stated that care should be taken in highway projects to ensure that organizations do not over-invest in conceptual estimating at the beginning phase. Since there is no guarantee, the project will be undertaken, easier and more effective ways to do conceptual cost estimating are highly recommended (Gardner et al. 2016).

The current approach towards construction cost estimating is based highly on the judgment of estimators because of the limited amount of information present at this stage of a project (Cheng et al 2009). As this might affect the estimation accuracy of the project, a scientific approach should be adopted for cost estimation at the conceptual stage (Cheng et al 2009). Various classical kriging methods are starting to be implemented to predict conceptual cost, however, they employ a single semi-variogram which is considered as the true semi-variogram, and the error associated with that is disregarded (Li et al. 2020). So, there is a need to find a kriging method which can solve this issue.

The objective of this study is to find a method that will give an accurate prediction of conceptual cost considering the location. The study is aimed at applying the empirical Bayesian kriging methodology together with three variogram models to estimate the conceptual cost of highway earthwork projects. After the semi-variograms are built, cross-validation will be done to check the errors and validate the models. The error variables between the different semi-variogram models will be compared and the best one will be chosen. The study will then be concluded by elaborating the contribution to knowledge and opening doors for further research on areas not covered.

Geostatistical modeling is a method used to predict variables whose values vary with location (Gupta et al. 2017). The process of estimating using spatial modeling involves building semi-variogram models which capture the spatial variation among variables (Gupta et al. 2017). The equation of the semi-variogram is given in equation 1 (Sağir and Kurtuluş 2017).

$$\gamma(h) = \tfrac{1}{2} * E[(Z(x) - Z^*(x - h))^2] \qquad (4.1)$$

where h = distance between the samples, Z(x) = the variable under study.

Kriging uses semi-variograms to predict values at unknown locations using the values at known locations (Mirzaei and Sakizadeh 2016). The study also emphasizes on how predicting errors are usually underestimated. In geostatistical kriging it is assumed that the data follows gaussian distribution (Basu 2016).

Empirical Bayesian kriging (EBK) is one type of geostatistical interpolation method (Mrazei and Sakizadeh 2016). EBK is a kind of estimating tool that uses a cluster of semi-variograms and allows the usage of moderately nonstationary data (Sağir and Kurtuluş 2017). Empirical Bayesian kriging uses more than one semi-variogram and it considers the error value associated with the semi variogram (Li et.al 2020). Semi-variogram parameters are plagued with uncertainties (Mrazei and Sakizadeh 2016). The method of EBK resolves that issue by creating a subset of data and drawing simulation which is a collection of multiple semi-variograms (Mrazei and Sakizadeh 2016). This method of kriging considers error variables (Mrazei and Sakizadeh 2016). Empirical Bayesian kriging is a kind of kriging that can work best both interactively and by automation which makes it rapid and dependable (Gribov and Krivoruchko 2020). This kriging type allows for interpolation on high numbers of data points up to billions of data (Gribov and Krivoruchko 2020).

According to the study by (Sağir and Kurtuluş 2017), there are three steps in drawing a model using EBK; First, a semi-variogram model will be built from the observed data, secondly, by using the semi-variogram created on the first step, a new value will be simulated at the same location of the observed data. Thirdly, using the simulated data, a new semi-variogram will be created at the same location. Finally, the method called the Bayes rule is used to measure the weight for the semi-variogram showing the likelihood of the development of original data from the semi-variogram. A similar process will then be going to be repeated after that (Sağir and Kurtuluş 2017). The study emphasizes that the parameters affecting the EBK should be specified, and the three important parameters presented are; (1) Subset size(2)Overlap factor and 3) Number of simulations; the method subdivides data into different subsets, and the number of data in each subset should be known; the subsets are not independent, they cross with each other and this parameter defines the degree of overlap between them ; each subset will have a certain number of semi-variograms that are going to be simulated and this factor specifies that number for each subset.

After the development of semi-variogram models, prediction of variables at the unknown location will be made using known points and the error associated will be measured (Gupta et al. 2017). Cross-validation is used to assess the effectiveness of the models and to compare the models, different types of errors will be calculated. The interpretations of error values used in this study were mean error, mean standardized error (MSDE), mean standard error (MSE), root-mean-square standardized error (RMSSDE), and root-mean-square error (RMSE) (Gupta et al. 2017). When mean error and mean standardized error approaches zero that is a representation of less error, value of Mean standard error (MSE) being close to the RMS is a measure of accurate prediction (Gupta et al. 2017). Root-mean-square standardized error (RMSSDE) is an error value which should be close to one for better prediction (Gupta et al. 2017). If it is greater than one, the prediction model underestimates the variability of the datasets; and if it is less than one, the prediction model overestimates the variability of the datasets (Gupta et al. 2017). Root-mean-square error (RMSE) Indicates how closely your model predicts the measured values (Gupta et al. 2017). The result for the cross validation will be used to answer research question three by choosing which combination of semi variogram with empirical Bayesian kriging gives better result with less error.

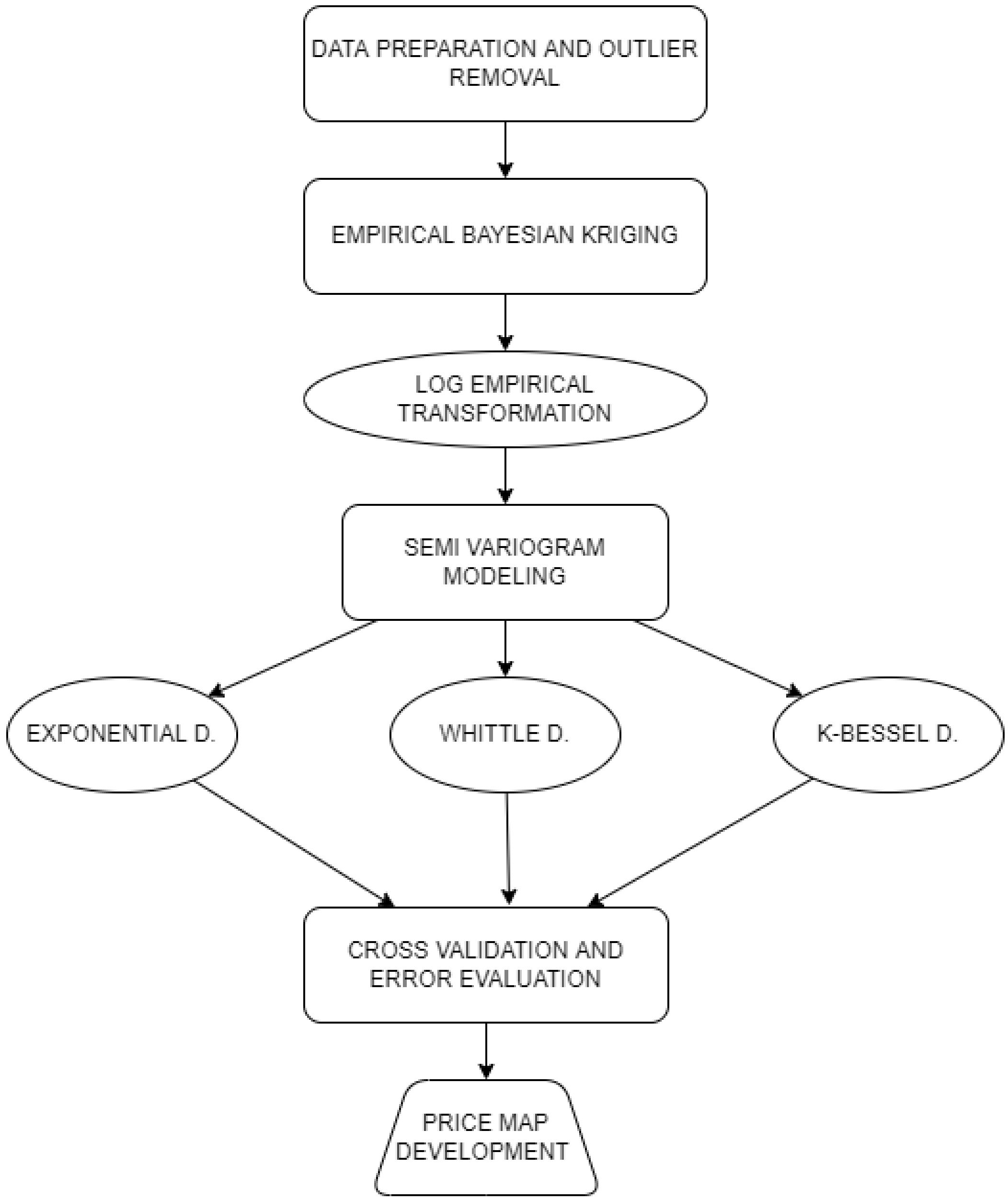


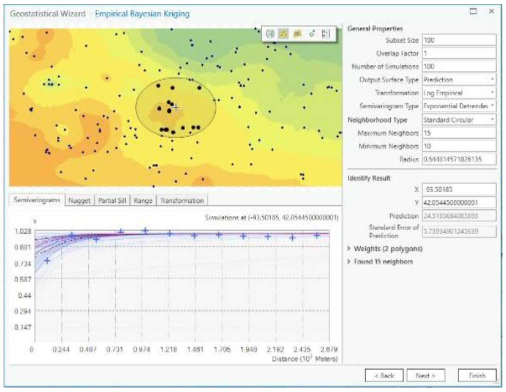
Fig. 4.1: Empirical Bayesian kriging approach in developing price model

Figure 1 shows the methodology followed by the research. Data for the study was collected from the Iowa Department of Transportation. Highway earthwork cost bid data for embankment, boulder excavation, and roadway and borrow excavation was collected for the year from 2015 to 2020 to predict the current price of 2021. The inflation factor was taken into consideration by extracting the price index values from the Iowa DOT website. Initial removal of outliers was done by careful analysis of histograms and QQ plots. The remaining outliers were found and removed using the Minitab software by using the Grubbs test technique. After the data was cleared for outliers, it was taken directly to the GIS. For locations where two or more sample exists at the same places, the maximum was used to handle coincident points.

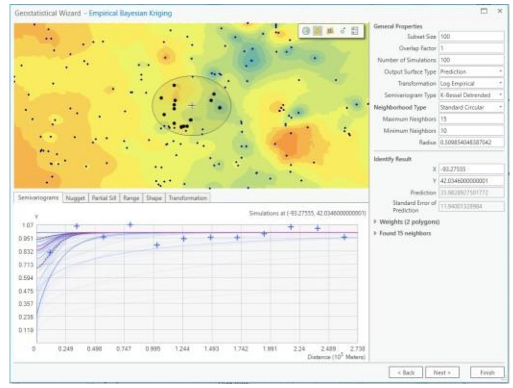
GIS gave the option of choosing transformation types before building the semi-variogram models. The options were, no transformation, empirical transformation, and log empirical transformation. For no transformation, three semi-variogram types were presented: linear, power, and thin-plate spline. For empirical and log-empirical transformations, the semi-variogram types available were exponential, exponential detrended, whittle, whittle detrended, k-Bessel, k-Bessel detrended. For this study log-empirical transformation was chosen and all semi variograms types were assessed. Since the majority of the detrended models show better results than the not detrended ones, they were chosen on most variables. But some variables showed better results with not detrended semi-variograms and those were also used. After the semi-variogram, the next step was applying the kriging interpolation followed by comparing the various types of errors using the cross-validation technique to validate the models. Then the cost map was drawn in the end. In this method not only one true semi-variogram was chosen and analyzed but a cluster of semi-variograms was analyzed. And the errors associated were carefully studied and presented.

4.2. Semi Variogram Estimation

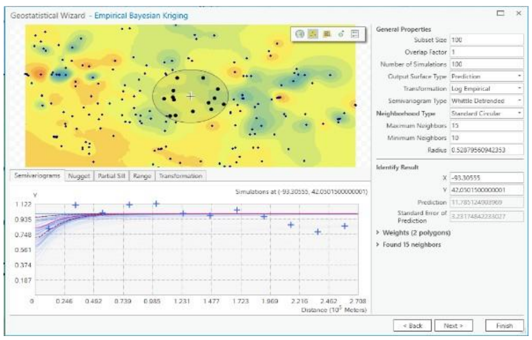
Log empirical transformation was implemented on all years from 2015 to 2020 on the three-highway earthwork bid items. The spectrum of semi-variograms is attached in the figures here. As shown below, the method of EBK is not represented by a single semi-variogram but a spectrum of semi variograms.



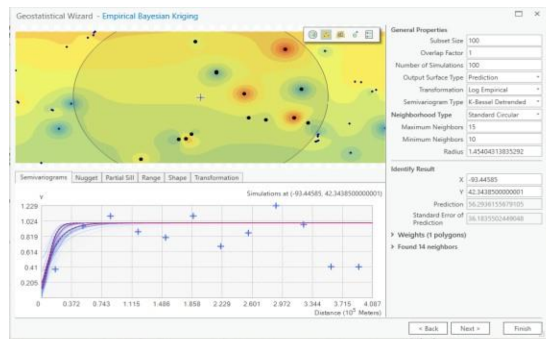
Log Empirical Exponential Detrended (Emb 2015)



Log Empirical K-Bessel Detrended (Emb 2017)



Log Empirical Whittle Detrended (Exc RB 2015)



Log Empirical K-Bessel Detrended (Exc Bou 2019)

Fig. 4.2: Semi variogram models of sample bid items-empirical Bayesian kriging

As shown in Fig 3, different types of semi-variogram models were used for the analysis; exponential detrended, whittle detrended, and k-Bessel detrended. The majority of the detrended models show better results than the not detrended ones so they were chosen on most. But still, some of them, the not detrended ones that gave good results with less error were also included. For the bid data Embankment 2016, a log empirical transformation with k-Bessel semi-variogram was implemented and it turned out the k-Bessel without the detrended semi-variogram seems to best fit the data. In the end, the k-Bessel semi-variogram of embankment 2016 was found to be better than the entire set. On embankment 2020 data also k-Bessel was included because it showed better results than the respective detrended one. For excavation boulder, all the exponential, whittle, and k-Bessel were included since they show better prediction than their respective detrended ones. Here again, the whittle model gave the best model than the whole set. On excavation RB 2016 all the exponential, whittle and k-Bessel were included owing to their better prediction than their respective detrended ones. Again, the whittle was shown to give a better result than the entire set. For Excavation RB 2018 again all the exponential, whittle, and k-Bessel were included owing to their better prediction than their respective detrended ones. Here also the whittle gave a better result than the rest. On excavation RB 2020, k-Bessel was included and, in this group, exponential detrended showed the better result.

4.3. Comparison of Models by Cross Validation

After the semi-variogram models were built, the validity of the models was checked using the parameters RMSE, ASE, MSDE, and RMSSDE. Table 1 shows sample data points used in the cross-validation process

Emb 2015

ERROR	Semi-Variogram Type with Log transformation			
	Exponential D.	Whittle D.		K-Bessel D.
Count	165	165		165
Average CRPS	8.49	8.49		8.49
Inside 90 % Interval	89.09	89.69		89.69
Inside 95 % Interval	95.15	95.15		95.15
Mean	0.33	0.27		0.45
Root-Mean Square	16.42	16.43		16.43
Mean Standardized	0.012	0.008		0.018
Root Mean Sq Standardized	0.94	0.93		0.92
Average Standard Error	17.69	17.89		18.11

EMB 2016

ERROR	Semi-Variogram Type with Log transformation			
	Exponential D.	Whittle D.	K-Bessel	K-Bessel D.
Count	183	183	183	183
Average CRPS	12.8	12.83	12.94	12.96
Inside 90 % Interval	92.35	92.35	91.26	91.8
Inside 95 % Interval	96.72	96.17	95.08	96.72
Mean	1.11	1.19	-0.304	1.65
Root-Mean Square	23.79	23.84	24.06	24.08
Mean Standardized	0.046	0.047	-0.012	0.056
Root Mean Sq Standardized	0.912	0.905	0.97	0.89
Average Standard Error	25.94	26.22	24.76	26.87

Exc RB 2018

ERROR		Semi-Variogram Type with Log transformation				
	Exponential	Exponential D.	Whittle	Whittle D.	K-Bessel	K-Bessel D.
Count	208	208	208	208	208	208
Average CRPS	4.01	4.01	4.02	4.02	4	4.02
Inside 90 % Interval	91.35	91.35	90.38	91.35	90.38	91.83
Inside 95 % Interval	95.19	95.19	95.19	95.19	95.19	94.71
Mean	-0.19	0.017	-0.245	0.12	-0.21	0
Root-Mean Square	7.37	7.37	7.37	7.39	7.35	7.4
Mean Standardized	-0.03	-0.005	-0.036	0.0085	-0.03	-0.01
Root Mean Sq Standardized	0.98	0.99	0.99	0.95	0.98	0.98
Average Standard Error	7.58	7.53	7.44	7.82	7.52	7.63

Fig. 4.3: Cross-validation results for sample cost data-empirical Bayesian kriging

According to the results from the cross-validation, from assessment of the error values: For embankment; 2015 exponential detrended showed the better result, 2016 k-Bessel, 2017 exponential detrended, 2018 whittle detrended, 2019 exponential detrended, 2020 exponential detrended showed better results. For excavation boulder, 2015 exponential detrended, 2016 k-Bessel detrended, 2017 whittle detrended, 2018- exponential detrended, 2019-whittle detrended, 2020-whittle showed better results. For excavation roadway and borrow; 2015 whittle detrended, 2016-whittle, 2017-exponential detrended, 2018-whittle, 2019-exponential detrended, 2020-exponential detrended showed better results. When comparing the entire sets, exponential detrended seems to show the better result in most years. Four years from embankment, two years from excavation boulder, and three years from excavation RB showed less error when exponential detrended was implemented which led to the decision of choosing exponential detrended semi variogram. In the end price map was developed. The figure below shows the price maps for sample cost data of embankment 2019 and excavation roadway and borrow 2018 and 2019.

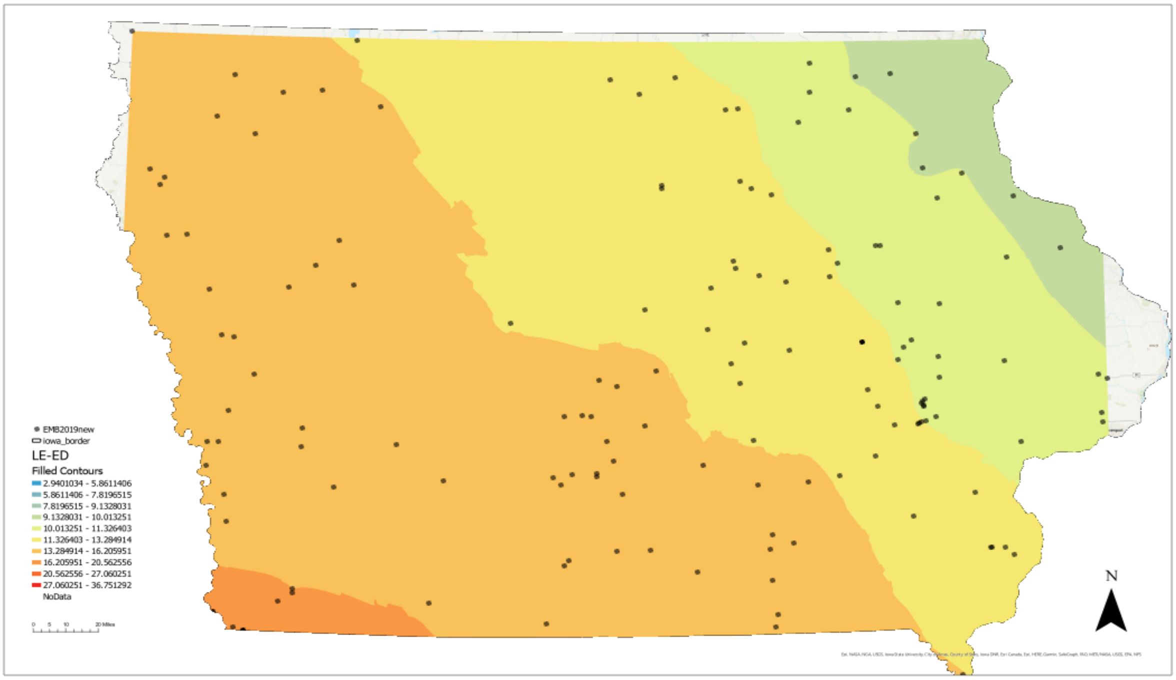


Fig. 4.4: Embankment 2019 price map

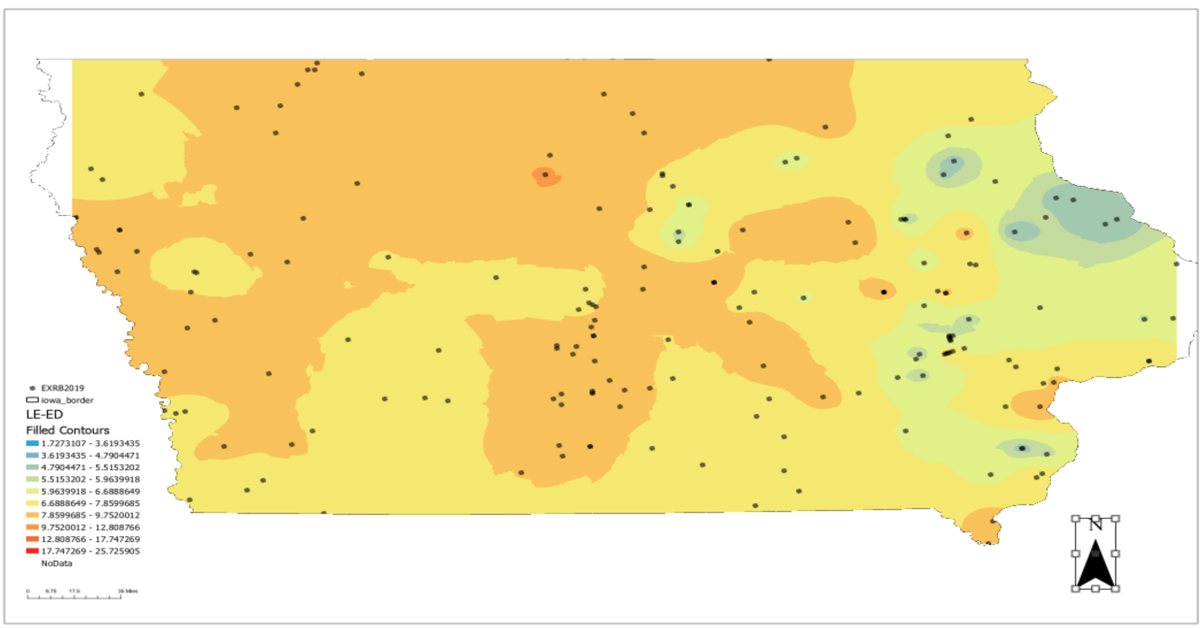


Fig. 4.5: Excavation roadway and borrow 2019 price map

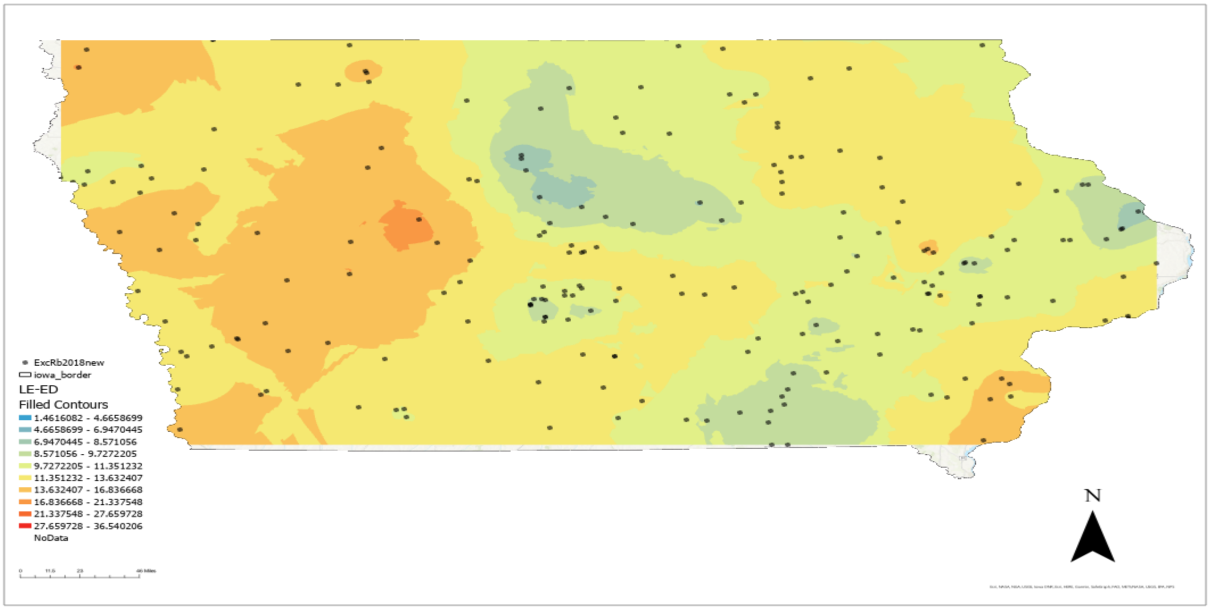


Fig. 4.6: Excavation roadway and borrow 2018 price map

4.4. Conclusion

It is challenging to estimate cost while the construction is at the initial stage. More specifically in highway earthwork construction, accurate conceptual cost estimation is very critical owing to its enormous amount taking the majority part of highway construction. Because of that, there is a need to find means of accurately estimating highway earthwork costs while the construction is in the beginning phase. An approach widely used in the construction industry is the judgment of experienced estimators using past cost data as a base. Human intuition is always prone to error and past data disregard the effect of time and location which is the reason for inaccurate estimation. Previous studies also tried to implement classical kriging methods which only consider a single semi-variogram and they disregard the error associated with that. Because of that, there is a need to find a scientific approach that considers the effect of location and that accounts for errors. The objective of this paper is to implement the empirical Bayesian kriging method using three different types of semi-variograms; exponential, whittle, and k-Bessel applying log empirical transformation on the data. Highway earthwork data of three representative bid items from the year 2015 to 2020 were collected from the Iowa department of transportation. To come up with the cost prediction for the current year 2021, the inflation factor was applied. A spectrum of semi-variogram models was used and errors were compared. The result showed that different semi-variograms gave a different result for each data but overall exponential detrended was shown to give a better result. This is a great contribution for the construction industry by implementing GIS into bid items to predict the cost. But in this study, the models were tested for only three representative bid items. In the future, this test could be expanded for more than the three bid items used.

4.5. References

Adel, K., Elyamany, A., Belal, A. M., & Kotb, A. S. (2016). Developing parametric model for conceptual cost estimate of highway projects. International Journal of Engineering Science, 6(7), 1728-1734.

Basu, K. (2016). Spatio temporal mapping of tuberculosis using log-empirical Bayesian kriging methods. Technical Report Indian Statistical Institute/Applied Statistics Unit.

Cheng, M. Y., Tsai, H. C., & Hsieh, W. S. (2009). Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. Automation in Construction, 18(2), 164-172.

Elbeltagi, E., Hosny, O., Abdel-Razek, R., & El-Fitory, A. (2014). Conceptual cost estimate of Libyan highway projects using artificial neural network. Int. J. Eng. Res. Appl, 4(8), 56-66.

Gardner, B. J., Gransberg, D. D., & Jeong, H. D. (2016). Reducing data-collection efforts for conceptual cost estimating at a highway agency. Journal of Construction Engineering and Management, 142(11), 04016057.

Gribov, A., & Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. Science of the Total Environment, 722, 137290.

Gupta, A., Kamble, T., & Machiwal, D. (2017). Comparison of ordinary and Bayesian kriging techniques in depicting rainfall variability in arid and semi-arid regions of north-west India. Environmental Earth Sciences, 76(15), 1-16.

Hyari, K. H., Al-Daraiseh, A., & El-Mashaleh, M. (2016). Conceptual cost estimation model for engineering services in public construction projects. Journal of Management in Engineering, 32(1), 04015021.

Kurtuluş, B., & Sağir, Ç. (2017). Hydraulic head and groundwater 111Cd content interpolations using empirical Bayesian kriging (EBK) and geo-adaptive neuro-fuzzy inference system (geo-ANFIS). Water SA, 43(3), 509-519.

Li, Y., Hernandez, J. H., Aviles, M., Knappett, P. S., Giardino, J. R., Miranda, R., Puy, M. J., Padilla, F., & Morales, J. (2020). Empirical Bayesian Kriging method to evaluate inter-annual water-table evolution in the Cuenca Alta del Río Laja aquifer, Guanajuato, México. Journal of Hydrology, 582, 124517.

Mirzaei, R., & Sakizadeh, M. (2016). Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran. Environmental Science and Pollution Research, 23(3), 2758-2769.

5. CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK UNIT PRICES: INDICATOR KRIGING APPROACH

5.1. Introduction

Finding accurate conceptual cost estimates of a highway earthwork project is a significant and challenging problem. Highway construction is a type of construction that takes several periods to be completed and involves complex construction process and the use of heavy equipment (Bogenberger et al. 2015). Road projects are basic ingredients in a country's development and a large amount of money is allocated for their construction (Choudhari et al. 2015). However, these projects are known to exceed the allotted time and cost (Choudhari et al. 2015). In highway construction, the first step involves the preparation of a survey and layout of the entire area where the construction is going to be performed (Bogenberger et al. 2015). After the completion of the layout design, earthwork construction is the task that follows in a highway construction activity (Bogenberger et al. 2015).

Earthwork is one of the main activities of highway construction which is complex and involves a high amount of cost (Bogenberger et al. 2015). Earthmoving is complex, expensive and involves adjusting and preparing subgrades of the proposed roadway (Choudhari et al. 2015). An enormous amount of earthwork is then placed on the proposed locations (Choudhari et al. 2015). The high amount of earth material being carried, and the long-haul distances are the reasons for earthwork to incur a high cost (Choudhari et al. 2015). During earthwork construction, a great amount of care must be taken owing to the influence it has on the surrounding environment (Bogenberger et al. 2015). The study added, since the impact earthwork makes on the surrounding is high, careful assessment technique is implemented on the structure and the environmental impact.

The main activities associated with earthwork are digging from one site with higher grades (cut areas) and moving, placing, and compacting materials to another site with lower grades (fill areas) to produce the final road layout (Bogenberger et al. 2015). Kim et al. (2015) also supports this by stating, earthwork is a very significant activity in highway construction, that involves cutting, moving, and filling of earth materials to bring the existing ground to the end road grade. Digging is a process of extracting and moving a vast amounts of earth materials using excavating equipment like excavators, whereas filling is a process of moving and placing earth materials in their respective fill areas, cuts, and trenches (Bogenberger et al. 2015). Creating the main components of a road by transporting, filling, and compacting large amounts of earthworks to the desired density involves the investment of a high amount of cost (Choudhari et al. 2015). The study emphasizes that earthwork not only involves digging and filling but also the management of the material being transported like obtaining, recycling, and disposing of it temporarily or permanently (Bogenberger et al. 2015). According to (Bogenberger et al. 2015), maintaining an optimal cost while performing the earthwork is a critical and challenging process.

The success of a certain construction project is dictated by how well the cost is accurately estimated when the project is at the design stage (Juszczyk 2013). Estimates made at the early stages have significant importance (Trost and Oberlender 2003). If cost is estimated over or under the actual, it will tremendously affect the success of a project (Juszczyk 2013). At the initial stage of a project, many parameters of the project are unknown, so assumption are used to will fill the gap in knowledge (Juszczyk 2013). Hence, estimates performed at this stage are used more on qualitative assumptions than actual quantitative ones (Juszczyk 2013). Despite it being usually decided upon general assumptions, the conceptual cost estimation performed at the beginning phase plays a significant role (Juszczyk 2013).

The accuracy of an estimate is a measurable phenomenon in which the accuracy is measured by the magnitude of how the predicted value approaches the accurate value (Oberlender and Trost 2001). The accuracy is governed by the following parameters: the individuals who prepare the estimation, how the estimation is prepared, and the known parameters while the estimation was performed (Oberlender and Trost 2001). Future cost estimates highly depend on early cost estimates using it as a base and the future estimate is expected to be beneath or equal to the early estimate, which is not usually the case (Oberlender and Trost 2001). The final estimate on most projects surpasses the early estimates because of the inaccuracy of the initial estimate due to limited scope definition at that stage (Oberlender and Trost 2001). A specific project is successful when the early estimate closely resembles the final estimate (Oberlender and Trost 2001).

Accurate early cost estimation is a very vital phenomenon for all stakeholders, specifically for the funding agency (Oberlender and Trost 2001). This is because all decisions that are involving the funding decisions, like allocations of assets and resources, will be based on the accuracy of early estimates (Oberlender and Trost 2001). Inaccuracy of early estimation could result in the loss of bids and project funds (Oberlender and Trost 2001).

Even though the early estimation is very critical, it is plagued by the limitation of information present at this stage, and this lack brings a significant number of errors to the estimation (Trost and Oberlender 2003). According to Trost and Oberlender (2003), estimating the cost at the conceptual stage has several constraints which contribute to its inaccuracy, namely a significant change in scope, limited time for preparation, difficulty to acquire complete information about cost data, change in design, and geography. Even though the early estimates are plagued with these uncertainties, it is one of the most important parameters utilized by owners and all stakeholders (Trost and Oberlender 2003).

This study is focused on estimating the conceptual cost of highway earthwork by using an indicator kriging together with exponential and k-Bessel semi-variograms. The study tried to compare the combination of the indicator kriging with the two semi-variograms to come up with the one which best estimates the cost. The highway earthwork data used for this study was acquired from the Iowa Department of transportation and it covers the years 2015 to 2020. All values were brought to the year 2021 by using cost index factors from the Iowa Department of Transportation. A geostatistical system called Geographic Information System (GIS) was used to do the analysis. The methodology part will explain the details of the research, and the outcome of the cross-validation results from the software will be explicitly explained in the following chapters.

5.2. Materials and Methods

Geographic Information System (GIS) was mainly utilized to perform the analysis of this study. The growth in usage of computer programs specifically computer graphics in different fields like urban, land management, and geocoding around 1960 and 1970, laid the foundation for the emergence of Geographic Information System (GIS) (Chang 2016). What makes GIS unique as compared to other business management systems is the fact that it incorporates geospatial data which is expressed in terms of location (Chang 2016). Geospatial data is data that is interpreted with regards to its location and attributes (Chang 2016). Road data is geospatial data defined by location and different attributes/ characteristics like length, name, speed limit, and direction (Chang 2016).

The component constituents of GIS are the users, geospatial data, the hardware, and the software (Chang 2016). GIS involves the process of obtaining, processing, and displaying data in a presentable form (Chang 2016).

A map can be prepared using GIS to present a report formally for a professional setting in such a way every element on the map like title and legend is written to make it more presentable (Chang 2016). On the other hand, maps can also be made informally with the sole purpose of manipulating the data on the software (Chang 2016). GIS has a feature that can assist in printing maps or saving it as a graphic file (Chang 2016). GIS has been used in a variety of disciplines since it came to existence; land use planning, wildlife habitat analysis, emergency planning, precision farming, and health are a few of most of the disciplines where GIS has found its application (Chang 2016). With the emergence and growth in usage of GIS, many researchers start showing interest in data acquisition, manipulation, storage, and display of spatial data (Cyril et al. 2019). GIS highly depend on several technologies for its advancement (Ali 2020). Computer programming, remote sensing, mathematics, and computer science are a few of the technologies which played significant roles in assisting the development of the software (Ali 2020).

Geostatistics is a branch of statistics that explains the spatial characteristics of data and its analysis by interpolation and mapping (Eldeiry and Garcia 2010). Geo statistics uses a probabilistic model to express the locational characteristics of the variable under study (Kurtulus and Sagir et al. 2017). Geo statistics incorporate statistical tools which allow coordinates expressing spatial locations of the data to be included (Tripathi et al. 2015). These spatial description assists in modeling spatial pattern of data, predicting values at locations where a sample is not taken, and assessing the errors that arose while predicting (Tripathi et al. 2015).

Geostatistics has a variety of functions (Eldeiry and Garcia 2011). Semi variogram is one of the main functions that assist the understanding of the spatial characteristics of data by the creation of a model which entails the spatial correlation among variables (Eldeiry and Garcia 2011). The variogram is a way in which locational characteristics will be presented by measuring the increase in variability of values as the distance between them increases (Kurtulus and Sagir et al. 2017). Building a semi-variogram model is the first step and it is a means by which the variability of random variables with location is computed (Liu et al. 2004). The characteristics of samples vary by location and these variograms are statistical tools that assist the study of that behavior (Asa et al. 2012). Dependency of one variable to the other in terms of location is the core concept of the semi-variogram model and it plots a graph in which y-axis is $\hat{\gamma}(h)$ and x-axis h distance between the points (Eldeiry and Garcia 2010).

Kriging model is a model that is used to predict values on locations where there are no samples taken (Eldeiry and Garcia 2010). When predicting using kriging, a weight is applied based on the nearest sample taken (Eldeiry and Garcia 2010). Autocorrelation measures and relates the kriging model to other interpolation methods (Johnston et al. 2001). It is a phenomenon by which things closer together will have similar characteristics whereas things further apart will differ (Johnston et al. 2001). What makes kriging like other methods specifically the Inverse distance weight (IDW) method is when predicting, kriging uses the weighted average of the surrounding points from which the measurement is taken (Johnston et al. 2001). In spatial autocorrelation, the weight is not only affected by location or how far the measured points are located, but also the general assembly between the points where the sample is taken (Johnston et al. 2001). What makes kriging like linear regression is the fact that both measure a weight that considers distance (Shamo et al. 2015). On the other hand, what makes kriging different is equal weight is allocated to reduce errors variance (Shamo et al. 2015).

Indicator kriging is a non-linear kriging type where the data values will be transformed with the binomial coding system using variables 0 and 1 based on a cutoff value zk (Glacken and Blackney 1998). Equation 1 presents the cut off values for a certain value z(x) (Glacken and Blackney 1998),

$$i(x; z_k) = \begin{cases} 1 & \text{if} \quad z(x) \geq z_k \\ 0 & \text{if} \quad z(x) < z_k \end{cases} \tag{5.1}$$

In indicator kriging, values that are slightly higher and which are significantly higher than a certain cut-off value (zk) will be given the same indicator value (Glacken and Blackney 1998). This made the method ideal because extremely high values will not have a significant effect on the result (Glacken and Blackney 1998). Indicator Kriging is a non-parametric geostatistical method whose approach of prediction of certain value at a certain point, is by estimating the probability by which a threshold value zk will be exceeded (Liu et. al 2004). In indicator kriging, the threshold/cutoff values are inclusive of the data scale which enables the data to be transformed into indicator values (Glacken and Blackney 1998). The semi-variogram will be modeled based on the transformed values (Glacken and Blackney 1998). The threshold defines the classification of the data by giving a value of 1 to all the data above the threshold and 0 to all the data below the threshold (Asa et al. 2012). Both for continuous and definite data, different covariances can be applied for different classes which are divided by thresholds (Ying 2000). This makes Indicator Kriging a better and more reliable method (Ying 2000). Since indicator kriging is non-parametric and it allows for mixed data populations, it became a preferable type of kriging widely used (Glacken and Blackney 1998). Indicator kriging is also famous for its application with highly skewed data sets (Glacken and Blackney 1998).

In geoscience, usually, data will have a highly skewed distribution (Armony 2005). Classical geostatistical disregard the shape of the distribution of the data, but indicator kriging works better with these kinds of skewed data (Armony 2005). Transforming the data into 0 and 1 made the coding for Indicator kriging simpler by putting the data together into a single process with common coding (Glacken and Blackney 1998). When estimating with multiple indicator kriging, it is necessary to define a specific number of thresholds that will be inclusive of the entire data set (Glacken and Blackney 1998). The study also states when choosing the number of thresholds, the time it takes to complete the analysis should also be considered. The existing problems with the present estimation methods; data not being distributed effectively with the boundaries and problems with parametric methods of analysis gave rise to the emergence of the indicator kriging (Glacken and Blackney 1998). Indicator kriging also found its usage in the mining field (Glacken and Blackney 1998).

Cross-validation assesses if the model appropriately fits the data (Berrar 2018). In cross-validation, after a value is estimated, the data will be replaced and a new one will be picked, then another round of estimation will be done on the remaining data (Liu et al. 2004). Cross-validation is a popular method of resampling in which model parameters can be manipulated and true errors associated with models can be predicted (Berrar 2018). In cross-validation, the performance will be estimated and tested quantitatively on new data while the model is being built (Berrar 2018)

The error values calculated by certain equations are significant indicators to say a model is acceptable or not (Tripathi et al. 2015).

Mean Error—is the deviation of the predicted values from the actual measured values and can be expressed as follows (Tripathi et al. 2015)

$$\mathrm{ME} = \sum_{i=1}^{n} \frac{[z^*(x) - z(x)]^2}{n} \quad (5.2)$$

Root Mean Square Error—Indicates the approach in which the model accurately predicts the measured value (Tripathi et al. 2015). A model is said to perform well when this value is smaller, and the equation is given below (Tripathi et al. 2015)

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{[z^*(x) - z(x)]^2}{n}} \quad (5.3)$$

Average Standard Error—is the mean value of the prediction standard error and has the following equation (Tripathi et al. 2015)

$$\mathrm{ASE} = \sqrt{\frac{\sum_{i=1}^{n} \sigma^*(x)}{n}} \quad (5.4)$$

Mean Standardized Error—is the average value of all the standardized errors which indicates the goodness of the model as it approaches zero and it's given in equation 7 (Tripathi et al. 2015)

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} \frac{z^*(x) - z(x)}{\sigma^*(x)}}{n} \quad (5.5)$$

Root Mean Square Standardized Error—is another way of testing a certain model which implies the goodness of a model as it approaches one (Tripathi et al. 2015). When this value is greater than one, the change in the estimation is said to be underestimated, whereas when this value is less than one the change is said to be overestimated and is represented by the equation as follows (Tripathi et al. 2015).

$$\mathrm{RMSSE} = \frac{\sum_{i=1}^{n} [\frac{\{z^*(x) - z(x)\}\hat{}2}{\sigma^*(x)}]\hat{}2}{n} \quad (5.6)$$

where z(x) represents the measured value, z*(x) equates the predicted value, and predicted value variance is represented by σ*(x) (Tripathi et al. 2015)

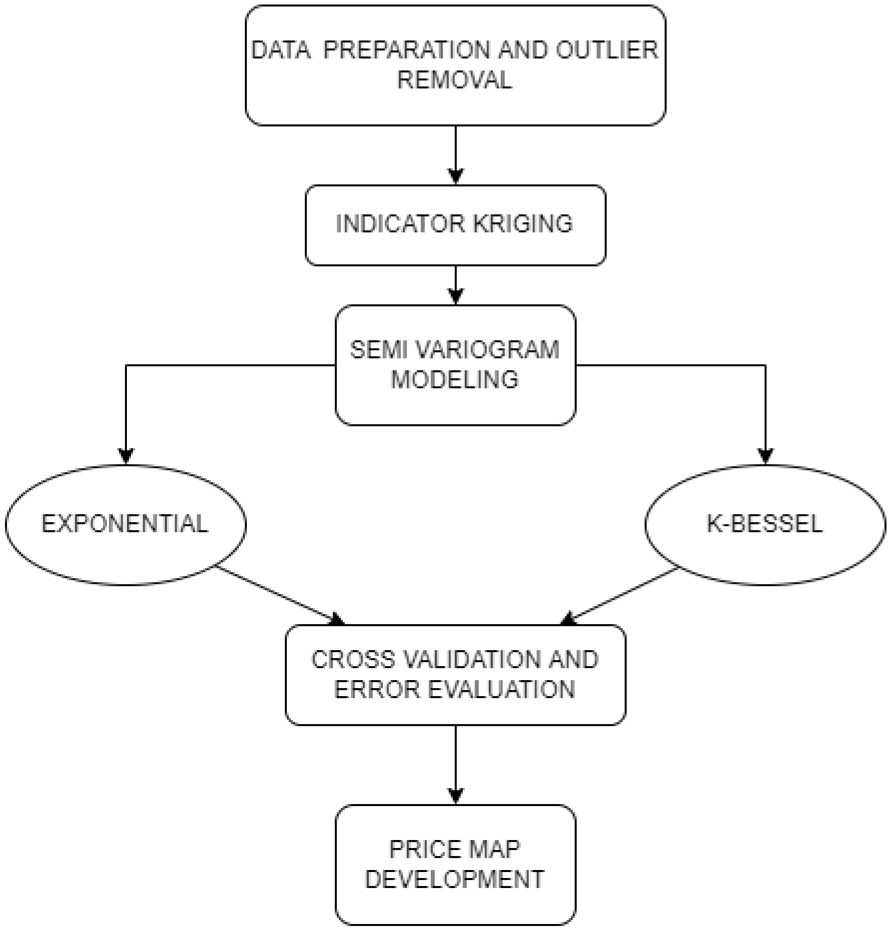


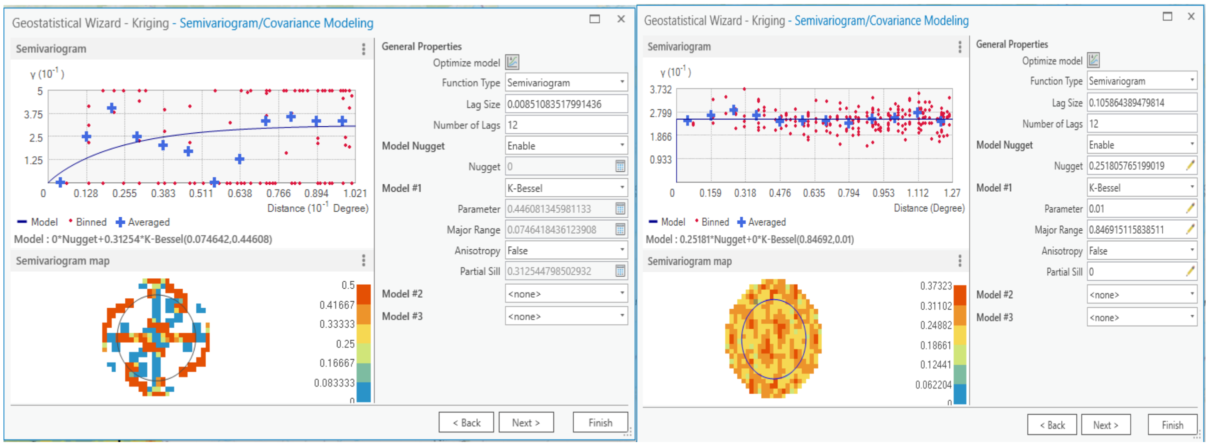
Fig. 5.1: Indicator kriging approach in developing price model

Figure 1 shows the method that the research followed to estimate the cost and create a price map. Iowa Department of transportation is the source of information in which the data was gathered for the study. Highway earthwork cost data was acquired from the Iowa Department of Transportation and the data for the years 2015 to 2020 were chosen for the analysis. The inflation factor extracted from the department website was applied to the data for each year to bring all the data to the year 2021. Outliers were removed from the data then the data was taken to GIS to conduct the analysis. In the beginning, it was found that two or more points were found to exist at the same location. For that, the maximum was chosen for the analysis.

GIS has a variety of kriging options to choose from ordinary, simple, universal, indicator, probability, and disjunctive kriging. The center of focus for the study which is indicator kriging was chosen. For the indicator kriging, map for probability and map for standard errors were the two options presented. The probability map was chosen since the study focuses on that. The mean value of the data was then used as a threshold value. For the threshold, the exceed option was chosen, which means values exceeding the threshold/mean will be given a value of one, and values less than the threshold will be given a value of zero. There was no additional number of cut-offs added, only one cut off, the mean value was used for this study. The next step was drawing the semi-variogram models. GIS gave a variety of models; Circular, spherical, tetra spherical, Penta spherical, exponential, gaussian, rational quadratic, hole effect, k-Bessel, J-Bessel, stable. For this study, exponential, and k-Bessel semi-variogram models were chosen. In the model building process, for function type the software gave two options; semi-variogram and covariance, semi-variogram was chosen. Model optimization was done for all functions. Model nugget was enabled, false anisotropy was chosen, and standard neighborhood type was used to build the model. After building the semi-variogram model, cross-validation was implemented to assess the performance of the model by carefully analyzing the errors. The error variables used to assess the performance of the models were mean, root-mean-square, mean standardized, root-mean-square-standardized, average standard error. These variables were assessed to measure the performance of the two models: exponential and k-Bessel.

5.3. Semi Variogram Modeling

Before the analysis has begun, outliers were detected and removed using Minitab and GIS. Grubbs technique on Minitab, and histogram and QQ plot charts on GIS were used to identify and clean the outliers. At first, indicator kriging paired with exponential semi-variogram was drawn for each year and each unit work separately and then followed by a K-Bessel variogram. The two models were then compared to see which performed better. When doing the analysis, all models were optimized to minimize errors. Two figures are attached here to show how the optimized model differs from the non-optimized for the unit work of excavation boulder for the year 2016 when the k-Bessel semi-variogram is used.



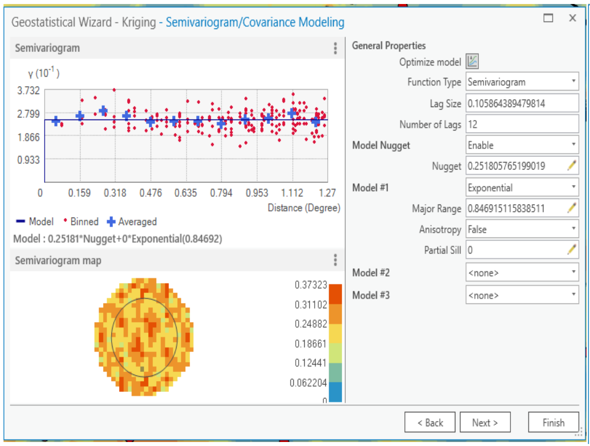
K-Bessel semi-variogram (Emb 2016)- before optimization

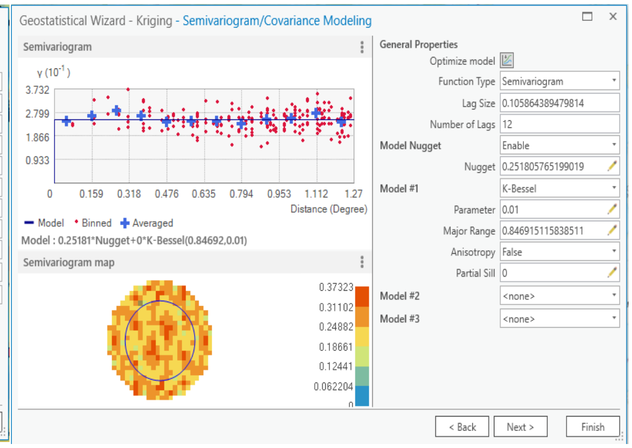
K-Bessel semi-variogram (Emb 2016)- after optimization

Fig. 5.2: K-Bessel semi-variograms before and after optimization-indicator kriging.

As it is shown in the comparison for the above two models, before optimization the data points were all highly spread away from the trend line. But for the same data when the model is optimized, the points seem to better fit the model because of the reduction of error of prediction. The same optimization of models was done for all items of work for all years.

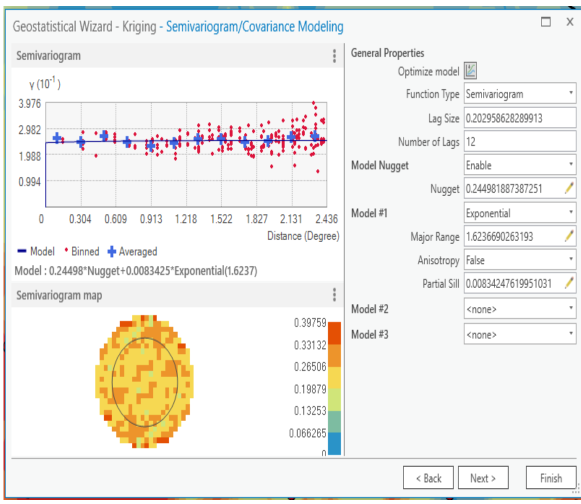
Attached here are sample semi-variograms drawn when the exponential and the k-Bessel variograms were implemented on each year and each unit item of work.

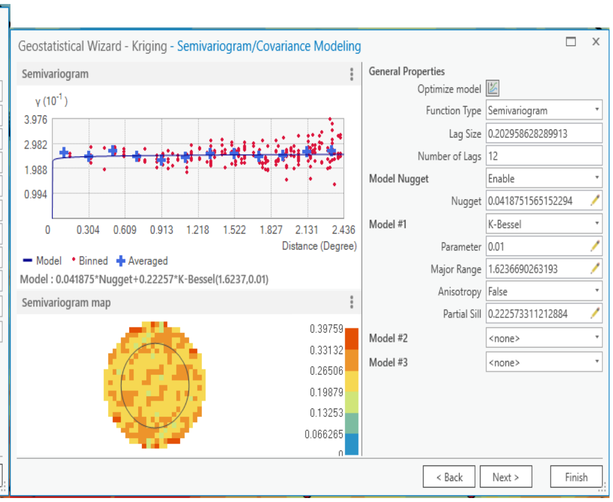




Optimized Exponential semi-variogram (Emb 2016)

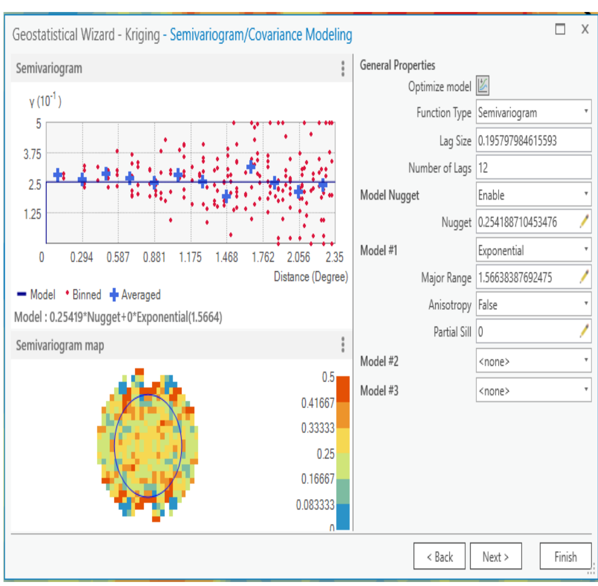
Optimized K-Bessel semi-variogram (Emb 2016)

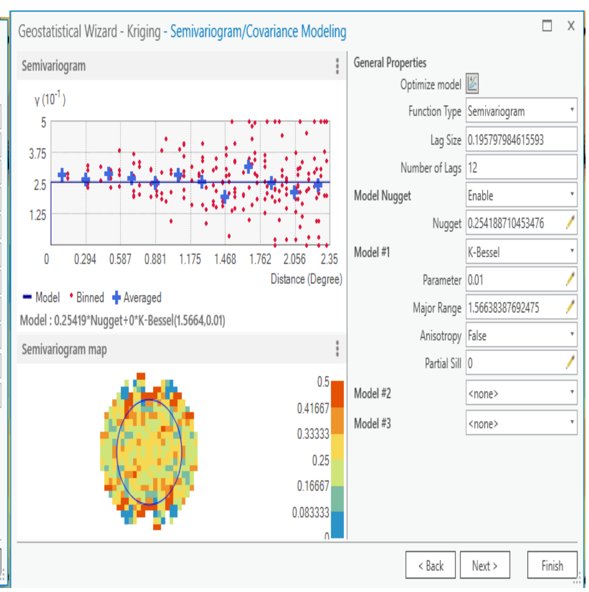




Optimized Exponential semi-variogram (Emb 2018)

Optimized K-Bessel semi-variogram (Emb 2018)

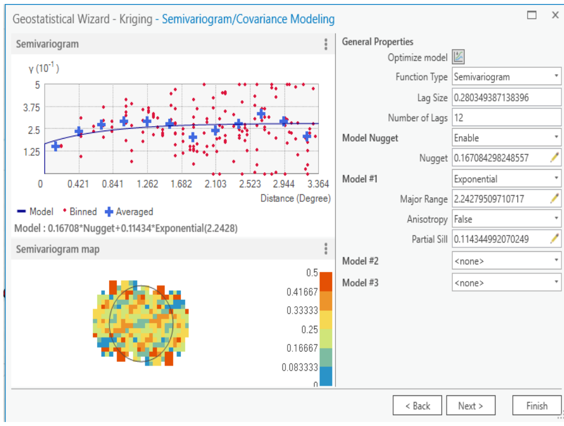




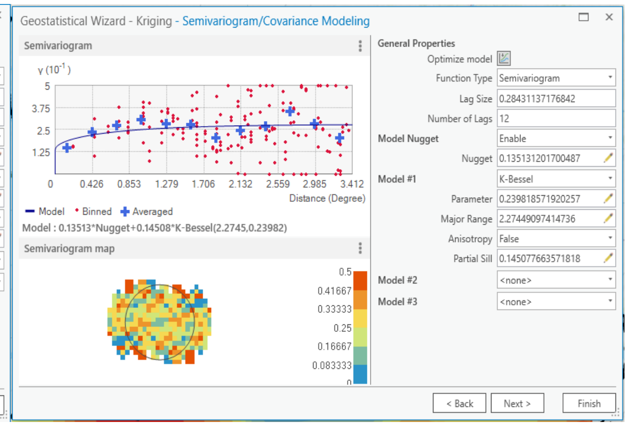
Optimized Exponential semi-variogram (Exc Bou 2016)

Optimized K-Bessel semi-variogram (Exc Bou 2016)

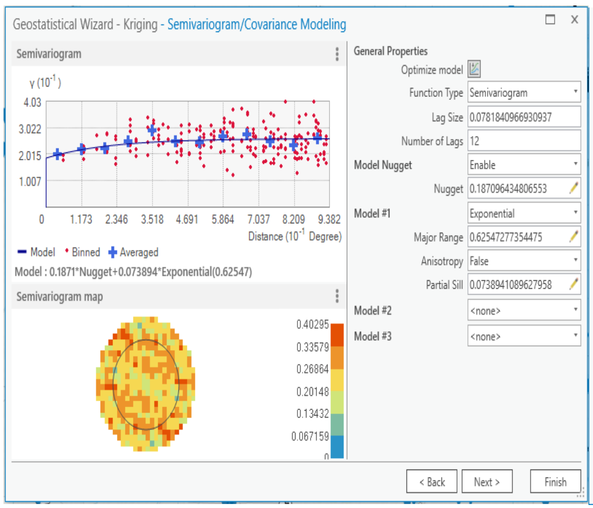
Fig. 5.3: Exponential and k-Bessel semi variograms for sample bid data-indicator kriging



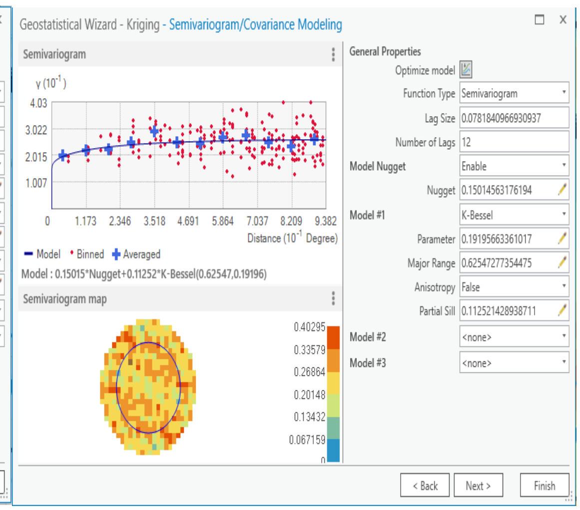
Optimized Exponential semi-variogram (Exc Bou 2018)



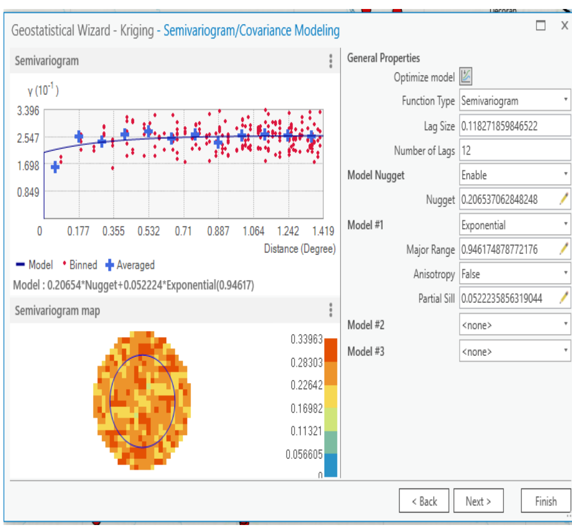
Optimized K-Bessel semi-variogram (Exc Bou 2018)



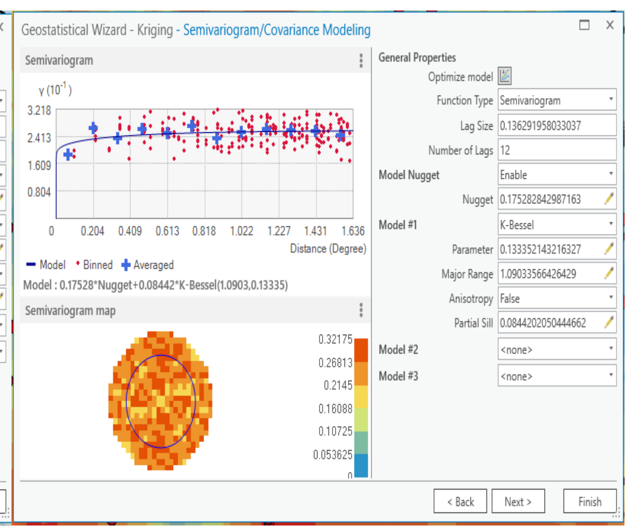
Optimized Exponential semi-variogram (Exc RB 2015)



Optimized K-Bessel semi-variogram (Exc RB 2015)



Optimized Exponential semi-variogram (Exc RB 2017)



Optimized K-Bessel semi-variogram (Exc RB 2017)

Fig. 5.3: Exponential and k-Bessel semi variograms for sample bid data-indicator kriging (continued)

As shown in the figures above, two semi-variograms; exponential and k-Bessel semi variograms for each unit work and each year were presented. All the semi-variograms were optimized. By careful observation of the semi variograms, it was found that Exponential semi-variogram seems to better fit the data

5.4. Cross Validation

An accurate way of validating would be analyzing the cross-validation results using error values. The error values that were used to decide on the validity of the models in this study are the mean and mean standardized, the root mean square standardized, the mean standardized and the average standard error. Attached here is the sample figure of the comparison between the two models.

EMB 2017		
ERROR	Semi-Variogram Type	
	Exponential	K-Bessel
Count	171	171
Mean	0.0179	0.0203
Root-Mean Square	0.51362	0.51261
Mean Standardized	0.03014	0.03902
Root Mean Sq Standardized	1.0383	1.12922
Average Standard Error	0.48868	0.44444
	0.02494	0.06817

EMB 2018		
ERROR	Semi-Variogram Type	
	Exponential	K-Bessel
Count	141	141
Mean	0.00733	0.00731
Root-Mean Square	0.50918	0.509
Mean Standardized	0.01371	0.01352
Root Mean Sq Standardized	0.99025	0.99443
Average Standard Error	0.51435	0.51204
	0.00517	0.00304

EXC BOU 2017		
ERROR	Semi-Variogram Type	
	Exponential	K-Bessel
Count	55	55
Mean	0.00054	0.00477
Root-Mean Square	0.47229	0.47179
Mean Standardized	0.00655	0.01268
Root Mean Sq Standardized	0.83507	0.85549
Average Standard Error	0.54788	0.52696
	-0.07559	-0.05517

EXC BOU 2018		
ERROR	Semi-Variogram Type	
	Exponential	K-Bessel
Count	44	44
Mean	0.00096	0.00021
Root-Mean Square	0.48306	0.48147
Mean Standardized	-0.00036	-0.00164
Root Mean Sq Standardized	0.98541	0.98428
Average Standard Error	0.48863	0.48786
	0.00557	0.00639

EXC RB 2017		
ERROR	Semi-Variogram Type	
	Exponential	K-Bessel
Count	202	202
Mean	-0.00306	-0.00338
Root-Mean Square	0.50635	0.50496
Mean Standardized	-0.0055	-0.006
Root Mean Sq Standardized	1.00845	1.00542
Average Standard Error	0.50178	0.50187
	0.00457	0.00309

EXC RB 2018		
ERROR	Semi-Variogram Type	
	Exponential	K-Bessel
Count	208	208
Mean	0.0048	0.00487
Root-Mean Square	0.49912	0.50178
Mean Standardized	0.01058	0.0106
Root Mean Sq Standardized	1.00324	1.01017
Average Standard Error	0.49717	0.49637
	-0.00195	-0.00541

Fig. 5.4: Cross-validation results for sample cost data-indicator kriging

The above-attached figures are two sample results for each item of work and year. The analysis for the other years and items of work was conducted and the results are presented as follows. According to the cross-validation result; For embankment 2015, the k-Bessel semi-variogram showed a better result. For excavation boulder 2015, the exponential semi-variogram showed a better result, and for excavation roadway and borrow 2015, the exponential semi-variogram performed better. From the assessment of the results for the year 2015, the overall exponential semi-variogram seems to perform better. For the year 2016, for both unit items of work, embankment, and excavation boulder, both k-Bessel and exponential showed similar results on all kinds of cross-validation error assessment results on the study. Since both show the same result, it is concluded that either one of the two models can be chosen. But for excavation roadway and borrow 2016, the exponential semi-variogram showed a better result. When summarizing, for 2016 overall exponential model was proven to perform better with less error. For the year 2017, the results are presented as follows. For embankment 2017, the exponential semi-variogram showed better result. For excavation boulder 2017, the k-Bessel semi-variogram showed a better result. For excavation RB 2017, the k-Bessel semi-variogram performed better which makes k-Bessel a better model for the year 2017. For the year of 2018; embankment 2018, k-Bessel semi variogram showed the better result, excavation boulder 2018, exponential semi-variogram showed the better result, and excavation RB 2018, exponential semi-variogram performed better which brings to the conclusion, for the year 2018, exponential-semi variogram seems to produce a better result. For the year of 2019; embankment 2019, exponential semi-variogram showed the better result, excavation boulder 2019, exponential semi-variogram showed better result, excavation roadway and borrow 2019, exponential semi-variogram performed better which makes exponentials semi-variogram a better model overall for 2019. For the year 2020; embankment, 2020 exponential semi-variogram showed a better result, excavation boulder, 2020 k-Bessel semi-variogram showed a better result, and excavation RB, 2020 k-Bessel semi-variogram performed better which makes k-Bessel variogram take the lead on the bid items understudy for the year 2020.

Summarizing the results for the different years; for 2015, exponential has shown better results of the model for unit items of work excavation boulder and excavation roadway and borrow, while the k-Bessel was better for the embankment unit work, which overall made the exponential take the lead. For the year 2016, for embankment and excavation boulder the two models showed similar results but for the excavation RB unit work, exponential showed the better result, which makes the exponential to perform better overall for that year. For the year 2017, exponential was a better model for the item of work embankment, but for the two items of work excavation boulder and excavation roadway and borrow, k-Bessel showed better result, so overall k-Bessel took the lead for 2017. For the year 2018, excavation boulder and excavation roadway and borrow showed better results for exponential, whereas embankment showed better results for k-Bessel, which makes exponential take the lead for this year. For the year 2019, on all items of work under study, the exponential showed the better result. For the year 2020, excavation boulder and excavation roadway and borrow showed better results for k-Bessel whereas embankment showed better results for exponential, making the k-Bessel to better predict the data for that year. From the entire result, it was seen that k-Bessel performed better with less error for the years 2017 and 2020 whereas all the other years' data performed better with the exponential semi-variogram, which makes exponential to be the chosen when the overall result is compared. The figure below shows the price maps for sample cost data of embankment 2019 and excavation roadway and borrow 2018 and 2019.

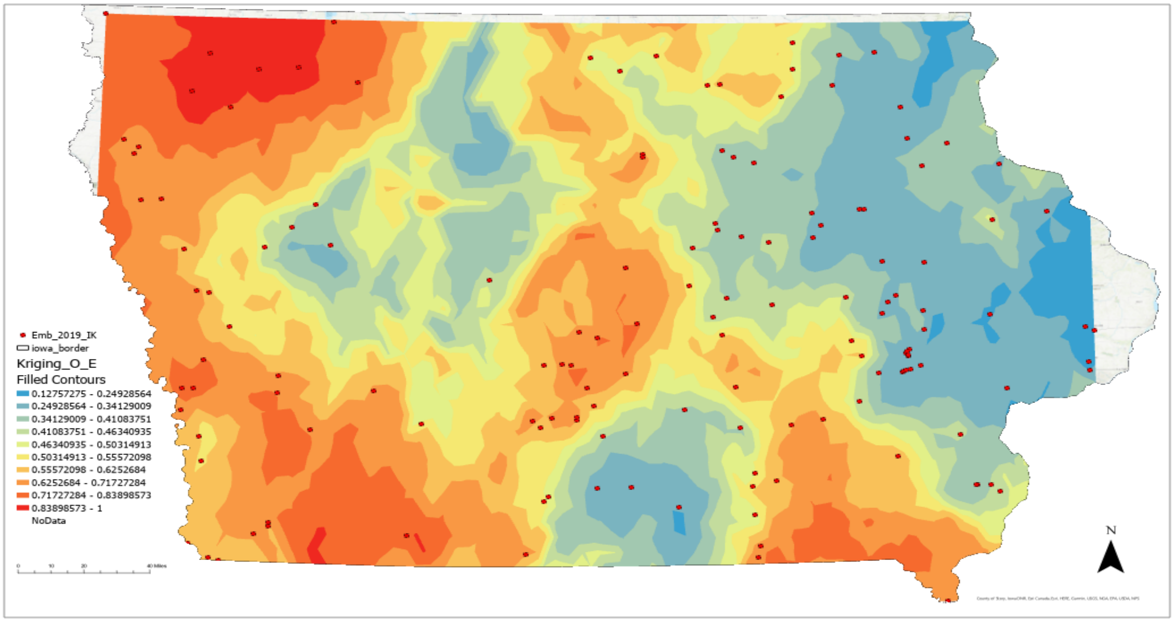


Fig. 5.5: Embankment 2019 price map

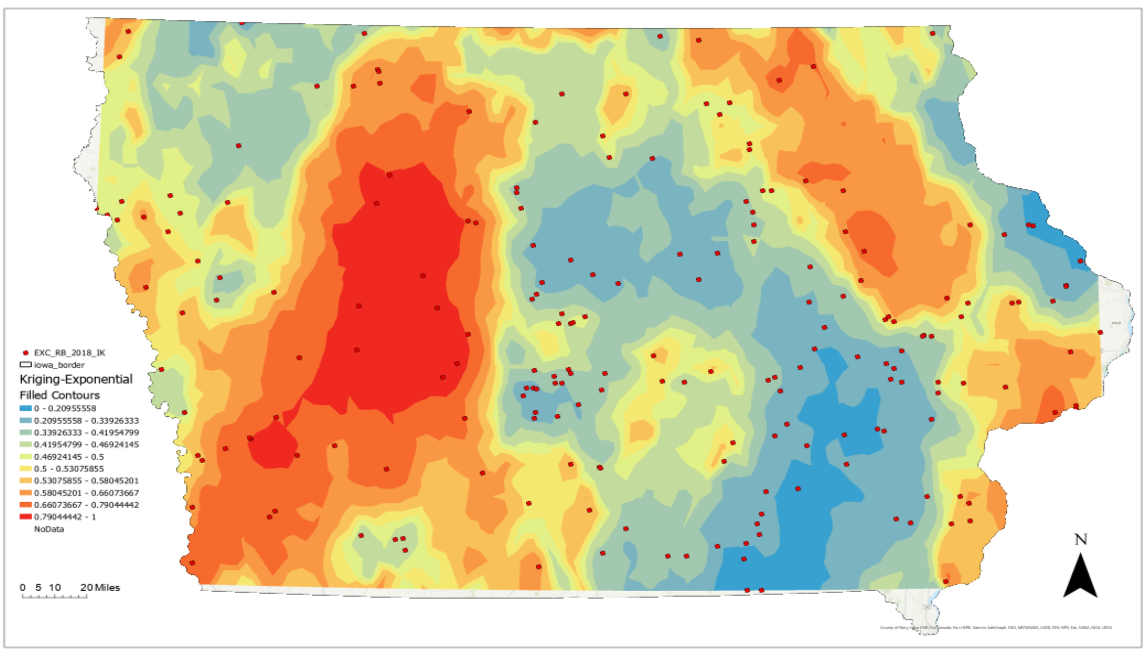


Fig. 5.6: Excavation roadway and borrow 2018 price map

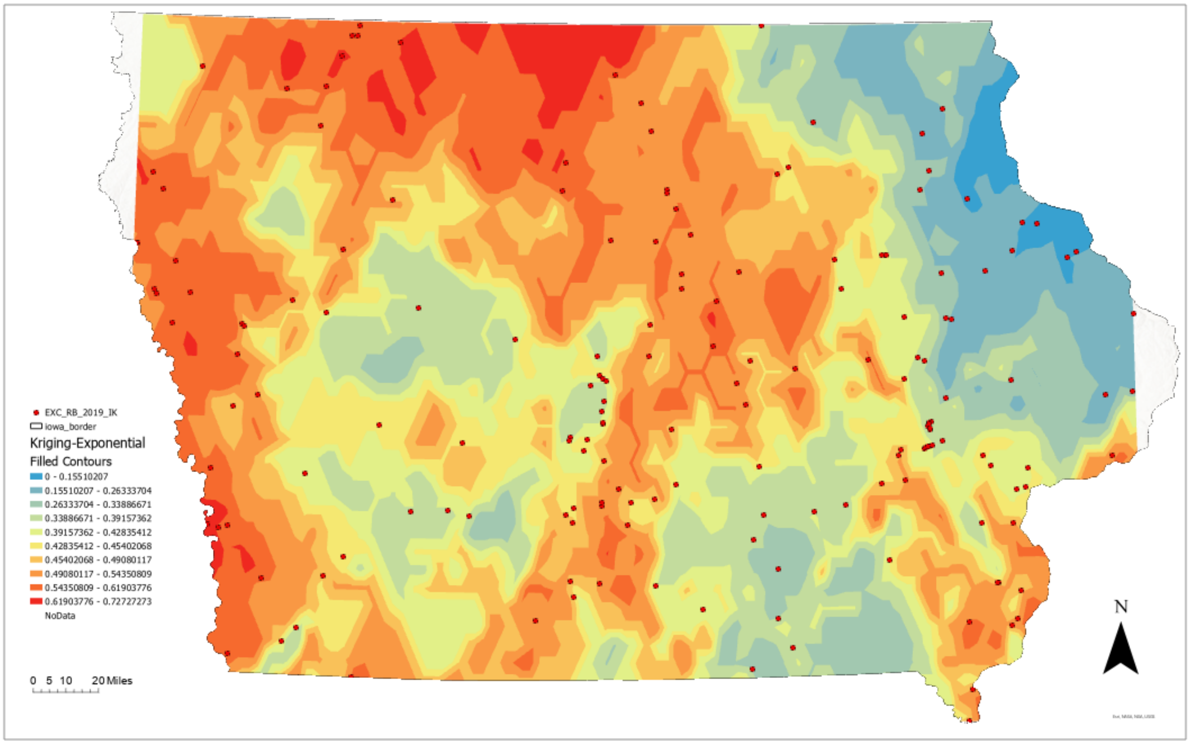


Fig. 5.7: Excavation roadway and borrow 2019 price map

5.5. Conclusion

It is challenging to prepare accurate estimations when the project is at the initial stage. Highway earthwork projects are known for their high magnitude of work and high investment involved which makes them even more detrimental if cost is estimated incorrectly at the beginning phase of a project. Highway earthwork is a significant unit item of work, but there is a gap in finding the best way to estimate the cost at the early stage because of the limited data available. This study was conducted to find a better estimating way for early cost estimation of highway earthwork by using past data from the year 2015 to 2020 from the Iowa department of transportation. The data for the years were brought to the year 2021 by using an inflation factor extracted from the department of transportation website. Two semi variograms: exponential and k-Bessel were tested to see which one performs better with indicator kriging. GIS was the medium used for the analysis. From the results obtained, it was noticed that the exponential semi-variogram performed better for the years 2015, 2016, 2018, and 2019 whereas the k-Bessel performed better for the years 2017 and 2020, which makes the exponential semi-variogram better for the specifically chosen unit items of work in this study. This study helps the earthwork construction industry by bringing a better conceptual cost estimation method. This study is specified only on three available bid items, so future studies can be made on more items of work and the results can be cross-checked.

5.6. References

Ali, E. (2020). Geographic Information System (GIS): Definition, Development, Applications & Components. Department of Geography, Ananda Chandra College. India.

Armony, M. (2005). Logical basis of Field Parametric Geostatistics. International Journal of Surface Mining, Reclamation and Environment, 19(2), 144-157.

Asa, E., Saafi, M., Membah, J., & Billa, A. (2012). Comparison of linear and nonlinear kriging methods for characterization and interpolation of soil data. Journal of Computing in Civil Engineering, 26(1), 11-18.

Berrar, D. (2018). Cross-validation. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 1–3(January 2018), 542–545

Bogenberger, C., Dell’Amico, M., Fuellerer, G., Hoefinger, G., Iori, M., Novellani, S., & Panicucci, B. (2015). Two-phase earthwork optimization model for highway construction. Journal of construction engineering and management, 141(6), 05015003.

Chang, K. T. (2016). Geographic information system. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology, 1-9.

Choudhari, S. C., Sahu, P. K., & Roy, D. (2015). Optimization of earthwork cost for road project using mathematical programming Optimization Model. 3rd Conference of Transportation Research Group of India, April 2018.

Cyril, A., Mulangi, R. H., & George, V. (2019). Development of a GIS-based composite public transport accessibility index. Journal of Urban & Environmental Engineering, 13(2).

Eldeiry, A. A., & Garcia, L. A. (2010). Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using LANDSAT images. Journal of Irrigation and Drainage Engineering, 136(6), 355-364.

Eldeiry, A. A., & Garcia, L. A. (2011). Using indicator kriging technique for soil salinity and yield management. Journal of irrigation and drainage engineering, 137(2), 82-93.

Glacken, I., & Blackney, P. (1998, October). A practitioner's implementation of indicator kriging. In Proceedings of a one-day symposium: Beyond Ordinary Kriging.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS geostatistical analyst (Vol. 380). Redlands: Esri.

Juszczyk, M. (2013, October). The use of artificial neural networks for residential buildings conceptual cost estimation. In AIP Conference Proceedings (Vol. 1558, No. 1, pp. 1302-1306). American Institute of Physics.

Kim, H., Chen, Z., Cho, C. S., Moon, H., Ju, K., & Choi, W. (2015). Integration of BIM and GIS: Highway cut and fill earthwork balancing. In Computing in Civil Engineering 2015 (pp. 468-474).

Kurtuluş, B., & Sağir, Ç. (2017). Hydraulic head and groundwater 111Cd content interpolations using empirical Bayesian kriging (EBK) and geo-adaptive neuro-fuzzy inference system (geo-ANFIS). Water SA, 43(3), 509-519.

Liu, C. W., Jang, C. S., & Liao, C. M. (2004). Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). Science of the Total Environment, 321(1-3), 173-188.

Oberlender, G. D., & Trost, S. M. (2001). Predicting accuracy of early cost estimates based on estimate quality. Journal of construction engineering and management, 127(3), 173-182.

Shamo, B., Asa, E., & Membah, J. (2015). Linear spatial interpolation and analysis of annual average daily traffic data. Journal of Computing in Civil Engineering, 29(1), 04014022.

Tripathi, R., Nayak, A. K., Shahid, M., Raja, R., Panda, B. B., Mohanty, S., Kumar, A., Lal, B., Gautam, P., & Sahoo, R. N. (2015). Characterizing spatial variability of soil properties in salt affected coastal India using geostatistics and kriging. Arabian Journal of Geosciences, 8(12), 10693-10703.

Trost, S. M., & Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. Journal of construction Engineering and Management, 129(2), 198-204.

Ying, Z. (2000). iksim: A fast algorithm for indicator kriging and simulation in the presence of inequality constraints, hard and soft data. Computers & Geosciences, 26(5), 493-507.

6. COMPARISON OF INDICATOR KRIGING AND EMPIRICAL BAYESIAN KRIGING FOR CONCEPTUAL COST ESTIMATION OF HIGHWAY EARTHWORK BID ITEMS

6.1. Introduction

The earthwork associated with road work is known for involving an enormous amount of materials that are moved from cut areas to fill areas; sometimes a long hauling distance (Askew et al. 2002). Earthwork is the major construction activity and involves cutting, transporting, filling, and compacting materials to bring the road level to the desired level along the route at considerable costs (Zitta et al. 2019). Minimizing the time and cost it takes to perform earthwork activity is fundamental to the successful completion of highway projects (Askew et al. 2002). The planning phase of earthwork is plagued with difficulties which are the results of the challenges associated with the selection of main earthwork activities (Askew et al. 2002). Decisions on the relationship between the activities and factors associated with the equipment used make the planning process difficult (Askew et al. 2002). Planning of earth materials must be done with extra care because of the high amount of cut and fill processes needed per a single strip of the road (Askew et al. 2002). Given the significant cost associated with an earthwork construction project, there is a need to develop accurate conceptual cost estimation (Zitta et al. 2019).

The planning and decision on the feasibility of a project are highly dependent on the accuracy of the estimate made at the initial phase of a project (Sodikov 2005). Estimating project cost at the early phase has its challenges owing to the limited amount of information, incomplete road work data, and absence of accurate cost estimation methods (Sodikov 2005). More information about the project will be known as the project proceeds which makes estimates made along the end more accurate than the estimates made at the initial phases of a project (Sodikov 2005).

According to Oberlender and Trost (2001), construction and engineering jobs benefit society the most when the early cost is estimated precisely. Accurate cost estimation assists project owners in making good budget decisions, saving them from the loss of opportunities (Oberlender and Trost 2001). Accurate cost estimation is also used by different project stakeholders as a measure of project success because project success is measured by how the early cost estimation and the actual cost after construction resembles (Oberlender and Trost 2001). Initial cost estimates are important because future estimates are based on them (Oberlender and Trost 2001). If initial cost estimates are done accurately, future estimates should be less than or equal to initial estimates (Oberlender and Trost 2001). But future estimates usually are higher than expected (Oberlender and Trost 2001). Even though the significance of the early cost estimate is higher, the information provided at the early stage is still limited (Oberlender and Trost 2001).

Given the importance of early conceptual cost estimation and the limited approach specifically on conceptual cost estimation of highway earthwork projects, this study employed Indicator kriging and empirical Bayesian kriging to estimate the cost of earthwork and compared the results of the two methods. Earthwork cost data set acquired from the Iowa department of transportation was used for the research. Data of an earthwork for three bid items embankment, excavation boulder, and excavation roadway & borrow from the year 2015 to 2020 were collected for the study. The error values from the cross-validation were used as a base of comparison. The methodology followed is presented in the coming chapters. A careful assessment of the results was done, and the study was concluded by suggesting the combination that gives a better result with less error.

6.2. Material and Methodology

Kriging types can be divided into two: linear and nonlinear; simple kriging, ordinary kriging, universal kriging, Bayesian kriging, and factorial kriging are listed as linear kriging whereas lognormal kriging, multi-gaussian kriging, disjunctive kriging, indicator kriging, probability kriging, and rank kriging are classified as nonlinear kriging (Asa et al. 2012).

There are some cases where it is impossible to use a gaussian random process for modeling certain kinds of distributions, where the data is positively skewed instead of being normally distributed (Moyeed and Papritz 2002). The study stated that in such cases, nonlinear kriging will perform better than linear. However, it should be used with extra care to ensure accuracy (Moyeed and Papritz 2002).

In this study, empirical Bayesian kriging and the nonlinear kriging type, indicator kriging was compared to estimate the unit price of highway earthwork projects. The data used for the study was collected from the Iowa department of technology of highway earthwork data from the year 2015-2020. The semi-variogram types used in the comparison are exponential and K-Bessel semi-variograms.

A variety of disciplines like engineering and earth sciences start to incorporate the principles of geostatistics for the prediction and simulation of various variables (Olea 2006). It needs important preparation and careful choices ahead of time by assessing the risks that might occur (Moyeed and Papritz 2002). What makes geo statistics unique as compared with the classical kinds of statistics is the analysis of spatial autocorrelations (Olea 2006). Geostatistics assume that if things are closer together, they will have similar values, and if they are further apart, they will have different values (Olea 2006). The study further elaborates that when there are smaller values, the values surrounding will also have smaller values and when there are large values, the values surrounding will also have larger values. Autocorrelation is considered as a core value in most applications of Geostatistics and it comes along with the development of semi-variogram and covariance (Olea 2006).

The development of a semi-variogram pertains to be the most significant characteristic of geostatistics application and many types of research exist around it (Olea 2006). Semi variogram is half the variance of the difference of two variables which are h distance apart (Olea 2006). When assessing the property of semi-variograms, there are always two points under the study, whose similarity in terms of separation distance and respective location are measured (Olea 2006). According to the study, uncertainty plays a significant role in statistics and geostatistics. The semi-variogram value calculated is an estimated value rather than being the actual true value which is unknown (Olea 2006). Among the different kinds of semi-variograms, the unbiased estimator one is the one that is popular, and it is shown in the equation below (Olea 2006).

$$\mathrm{Y(h)} = 1/2n(h) \sum_{i=1}^{n(h)} [\mathrm{z(Xi + h) - z(Xi)}]\text{^}2 \tag{6.1}$$

Where: Y(h) is experimental semi-variogram

Xi is the location of a point

Z(Xi) is the measurement at location Xi

h is the distance between points or lag

n(h) is the number of pairs h units apart in the direction of the vector

Some of the most common types of simple semi-variogram models are presented here (Olea 2006)

For $0<\partial$ and $0<C$

$$\text{Exponential Model: } Ex(h) = C\left(1 - e^{-3h/a}\right) \tag{6.2}$$

$$\text{Gaussian Model: } G(h) = C\left(1 - e^{-3(\frac{h}{a})^2}\right) \tag{6.3}$$

$$\text{Power Model: } P(h) = \partial h^{ß}, 0 < ß < 2 \tag{6.4}$$

$$\text{Sine hole effect: } S(h) = C\left(1 - \frac{\sin\left(\frac{\pi h}{a}\right)}{\frac{\pi h}{a}}\right) \tag{6.5}$$

When performing any kind of kriging it is necessary to model a semi-variogram (Olea 2006). It has always been an issue finding weights to minimize error which gave rise to the term called quadratic minimization error in kriging (Olea 2006). According to (Krivoruchko 2001) kriging is known for interpolating exact points. At points where data is not collected, kriging prediction will have a smooth change in the prediction map but when it reaches the points where data is collected, it will show a sprint to align to the measured value (Krivoruchko 2001). The significant aspect of the kriging method is it considers location (Luo et al. 2017). The study further stated errors of predicted values will be minimized and accuracy will be maintained by kriging. Indicator kriging draws probability maps which divides the probability of occurrence of values by a certain threshold (Krivorishko 2001). According to (Lloyd and Atkinson 2001), indicator kriging brings a solution to the normality requirements of a data and indicates the independence of standard errors on the data. The same study states that in indicator kriging different cut-off values are taken and at each cut-off value the conditional cumulative distribution function (CCDF) is predicted by the least square estimate (Lloyd and Atkinson 2001).

Geographic Information system (GIS) was the system used to perform the analysis. Geographic is to state the fact that location is a key factor, Information is to emphasize the importance of the information that will be gained from the data analysis to decide, and the system is to state that the whole user, hardware, and software is working as a system to give information (Ali 2020). GIS is a computer-based program that helps in making decisions by manipulating and presenting data that are referenced based on their location (Ali 2020). Geographic information is generated, stored, processed, and presented by using the computing software called GIS which incorporates software, hardware, data, users, and institution (Ali 2020).

Cross-validation is a means that compares interpolated and actual values (Asa et al. 2012). After the prediction is done using kriging, true prediction error which relates the actual(Z*(xi)) with the estimated(Z(xi)) will be calculated (Asa et al. 2012). The error equations are presented as follows (Asa et al. 2012)

$$\text{Mean Prediction Error (Mean)} = 1/n \sum_{i=1}^{n}[z(xi) - z^{\wedge} * (xi)] \qquad (6.6)$$

$$\text{Standard Mean Prediction Error (SM)} = 1/n \sum_{i=1}^{n}(Mean/\delta^{\wedge}2(xi)) \qquad (6.7)$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{1}/n \sum_{i=1}^{n}[z(xi) - z^{\wedge} * (xi)]\ {}^{\wedge}2 \qquad (6.8)$$

$$\text{Standardized root mean square prediction error (RMSES)} = \sqrt{1}/n \sum_{i=1}^{n}(Mean/\delta^{\wedge}2(xi))\ {}^{\wedge}2) \qquad (6.9)$$

$$\text{The average standard error (ASE)} = S/\sqrt{n} \qquad (6.10)$$

Where s = standard deviation

n = number of data

$\delta^{\wedge}2(xi)$ = variance at location i (Asa et al. 2012),

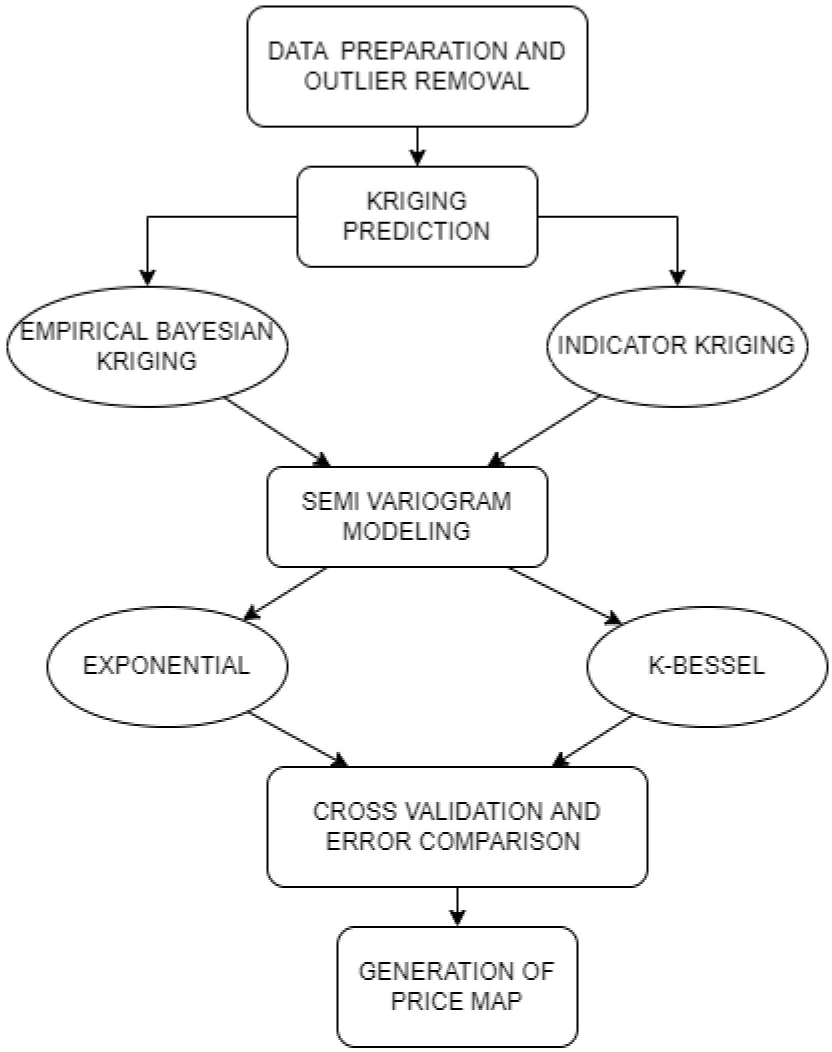


Fig. 6.1: Research methodology

Fig 1 depicts the methodology the study is going to follow. Data for embankment, for excavation boulder, and excavation roadway & borrow from the year 2015 to 2020 were gathered from the Iowa Department of Transportation. Outliers were removed using GIS and Minitab. The inflation cost index to bring the prices to the year 2021 was applied on all the data points from 2015 to 2020. For places where two or more samples exist at the same location, the maximum was chosen. The complete data set was then tested both for empirical Bayesian kriging and indicator kriging. For the EBK, log transformation was chosen, and exponential and k-Bessel semi-variograms were selected. For the Indicator, the threshold was chosen as exceeding, numbers exceeding the threshold were given a value of one, and numbers below were given a value of zero. The mean was chosen to be the threshold value, and no additional cutoffs were taken for the analysis. The models were all optimized and similarly exponential and k-Bessel semi-variograms were chosen to aid the comparison. After building the semi-variogram, the next step followed was cross-validating and comparing the error value to check which of the models performed better. Mean, Root-mean-square, mean standardized, root mean square standardized, and average standard error were the error parameters that were used to do the comparison. The cross-validation contributed to the final decision regarding the choice of the model.

6.3. Cross Validation Results

Error-values were compared, and models were validated by carefully analyzing if the root mean square and average standard error are close enough, the root mean square standardized is close to one or if the mean and mean standardized are close to zero.

EMB 2015				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	165	165	165	165
Average CRPS	8.488		8.496	
Inside 90 % Interval	90.303		90.909	
Inside 95 % Interval	95.152		95.757	
Mean	0.324	-0.00016	0.334	-0.00031
Root-Mean Square	16.415	0.49577	16.415	0.49637
Mean Standardized	0.016	-0.00043	0.016	-0.0007
Root Mean Sq Standardized	0.928	0.97419	0.919	0.979
Average Standard Error	17.699	0.50896	17.877	0.50699
	1.284	0.01319	1.462	0.01062

EMB 2016				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	183	183	183	183
Average CRPS	12.946		12.944	
Inside 90 % Interval	91.803		91.257	
Inside 95 % Interval	95.082		95.082	
Mean	-0.337	0.00097	-0.304	0.00097
Root-Mean Square	24.06	0.51541	24.056	0.51541
Mean Standardized	-0.013	0.00181	-0.012	0.00181
Root Mean Sq Standardized	0.972	0.99479	0.968	0.99479
Average Standard Error	24.636	0.51851	24.763	0.51851
	0.576	0.0031	0.707	0.0031

EMB 2017				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	171	171	171	171
Average CRPS	15.731		15.782	
Inside 90 % Interval	89.474		88.888	
Inside 95 % Interval	93.567		93.567	
Mean	-0.898	0.0179	-0.368	0.0203
Root-Mean Square	30.463	0.51362	30.573	0.51261
Mean Standardized	-0.04	0.03014	-0.024	0.03902
Root Mean Sq Standardized	1.053	1.0383	1.023	1.12922
Average Standard Error	29.358	0.48868	30.47	0.44444
	1.105	0.02494	0.103	0.06817

EMB 2018				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	141	141	141	141
Average CRPS	7.057		7.059	
Inside 90 % Interval	92.198		92.198	
Inside 95 % Interval	95.744		95.745	
Mean	-0.346	0.00733	-0.234	0.00731
Root-Mean Square	12.685	0.50918	12.687	0.509
Mean Standardized	-0.025	0.01371	-0.017	0.01352
Root Mean Sq Standardized	0.967	0.99025	0.968	0.99443
Average Standard Error	13.09	0.51435	13.127	0.51204
	0.405	0.00517	0.44	0.00304

EMB 2019				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	146	146	146	146
Average CRPS	3.757		3.774	
Inside 90 % Interval	86.986		86.986	
Inside 95 % Interval	92.466		93.15	
Mean	-0.095	-0.0112	-0.086	-0.00679
Root-Mean Square	6.939	0.50192	6.962	0.50175
Mean Standardized	-0.017	-0.02085	-0.019	-0.01263
Root Mean Sq Standardized	0.986	1.0073	0.99	1.0236
Average Standard Error	7.094	0.49897	7.129	0.49032
	0.155	-0.00295	0.167	-0.01143

EMB 2020				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	167		167	
Average CRPS	8.147		8.157	
Inside 90 % Interval	89.82		89.82	
Inside 95 % Interval	94.611		94.61	
Mean	0.39	0.00088	0.337	0.00059
Root-Mean Square	15.613	0.51715	15.623	0.51656
Mean Standardized	0.018	0.00252	0.014	0.00218
Root Mean Sq Standardized	0.932	1.01305	0.932	1.01391
Average Standard Error	17.043	0.51047	17.098	0.50943
	1.43	-0.00668	1.475	-0.00713

Fig. 6.2: Cross-validation result for embankment-comparison

EXC. BOU 2015				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	59	59	59	59
Average CRPS	21.965		21.841	
Inside 90 % Interval	94.915		94.915	
Inside 95 % Interval	96.61		96.61	
Mean	1.849	0.03386	0.799	0.03179
Root-Mean Square	39.937	0.48702	39.772	0.47893
Mean Standardized	0.017	0.05972	-0.004	0.05495
Root Mean Sq Standardized	0.83	0.88392	0.883	0.85896
Average Standard Error	51.807	0.55303	50.691	0.56232
	11.87	0.06601	10.919	0.08339

EXC. BOU 2016				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	53	53	53	53
Average CRPS	54.928		55.015	
Inside 90 % Interval	90.566		90.566	
Inside 95 % Interval	100		100	
Mean	4.567	0.00779	4.449	0.00779
Root-Mean Square	105.246	0.52969	105.397	0.52969
Mean Standardized	0.031	0.01368	0.029	0.01368
Root Mean Sq Standardized	0.787	1.01161	0.796	1.01161
Average Standard Error	134.281	0.52374	133.376	0.52374
	29.035	-0.00595	27.979	-0.00595

EXC. BOU 2017				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	55	55	55	55
Average CRPS	33.256		29.773	
Inside 90 % Interval	94.545		94.545	
Inside 95 % Interval	98.182		98.182	
Mean	0.931	0.00054	4.66	0.00477
Root-Mean Square	63.653	0.47229	58.007	0.47179
Mean Standardized	0.02	0.00655	0.022	0.01268
Root Mean Sq Standardized	0.834	0.83507	0.865	0.85549
Average Standard Error	78.639	0.54788	82.697	0.52696
	14.986	-0.07559	24.69	-0.05517

EXC. BOU 2018				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	44	44	44	44
Average CRPS	31.302		30.505	
Inside 90 % Interval	88.636		90.909	
Inside 95 % Interval	95.454		95.454	
Mean	1.835	0.00096	2.163	0.00021
Root-Mean Square	61.766	0.48306	60.305	0.48147
Mean Standardized	-0.002	-0.00036	-0.009	-0.00164
Root Mean Sq Standardized	0.956	0.98541	0.988	0.98428
Average Standard Error	73.534	0.48863	69.428	0.48786
	11.768	0.00557	9.123	0.00639

EXC. BOU 2019				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	44	44	44	44
Average CRPS	18.539		18.461	
Inside 90 % Interval	79.545		79.545	
Inside 95 % Interval	93.182		95.454	
Mean	0.509	-0.00571	0.873	-0.00875
Root-Mean Square	35.359	0.48324	35.233	0.48854
Mean Standardized	-0.038	-0.01181	-0.029	-0.01637
Root Mean Sq Standardized	1.137	1.04804	1.079	1.0845
Average Standard Error	31.809	0.45765	33.89	0.44342
	3.55	-0.02559	1.343	-0.04512

EXC. BOU 2020				
ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	45	45	45	45
Average CRPS	29.521		28.677	
Inside 90 % Interval	91.111		91.111	
Inside 95 % Interval	100		100	
Mean	2.402	-0.00137	3.454	0.00077
Root-Mean Square	56.215	0.44197	55.029	0.4458
Mean Standardized	0.025	-0.00045	0.029	-0.00257
Root Mean Sq Standardized	0.859	1.01596	0.831	0.87806
Average Standard Error	66.509	0.40746	65.872	0.46619
	10.294	-0.03451	10.843	0.02039

Fig. 6.3: Cross-validation result for excavation boulder-comparison

EXC. RB 2015

ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	195	195	195	195
Average CRPS	4.408		4.342	
Inside 90 % Interval	91.282		91.282	
Inside 95 % Interval	94.871		94.871	
Mean	0.056	0.00186	0.092	0.00313
Root-Mean Square	8.465	0.50827	8.356	0.50709
Mean Standardized	0.006	0.00195	0.014	0.00379
Root Mean Sq Standardized	0.978	1.00519	0.98	1.00428
Average Standard Error	8.715	0.50485	8.569	0.50404
	0.25	-0.00342	0.213	-0.00305

EXC. RB 2016

ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	226	226	226	226
Average CRPS	7.27		7.31	
Inside 90 % Interval	88.496		87.168	
Inside 95 % Interval	94.69		95.133	
Mean	-0.582	-0.00083	-0.473	-0.00054
Root-Mean Square	13.834	0.48304	13.903	0.48143
Mean Standardized	-0.043	-0.00417	-0.0396	-0.00438
Root Mean Sq Standardized	0.964	1.05743	0.978	1.19144
Average Standard Error	14.613	0.45084	14.564	0.39385
	0.779	-0.0322	0.661	-0.08758

EXC. RB 2017

ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	202	202	202	202
Average CRPS	7.343		7.304	
Inside 90 % Interval	89.604		89.109	
Inside 95 % Interval	97.029		97.029	
Mean	0.781	-0.00306	0.592	-0.00338
Root-Mean Square	14.325	0.50635	14.259	0.50496
Mean Standardized	0.048	-0.0055	0.038	-0.006
Root Mean Sq Standardized	0.863	1.00845	0.876	1.00542
Average Standard Error	16.585	0.50178	16.204	0.50187
	2.26	0.00457	1.945	0.00309

EXC. RB 2018

ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	208	208	208	208
Average CRPS	4.013		4.002	
Inside 90 % Interval	91.346		90.385	
Inside 95 % Interval	95.192		95.192	
Mean	-0.191	0.0048	-0.21	0.00487
Root-Mean Square	7.368	0.49913	7.353	0.50178
Mean Standardized	-0.027	0.0106	-0.03	0.0106
Root Mean Sq Standardized	0.976	1.00324	0.984	1.01017
Average Standard Error	7.585	0.49717	7.516	0.49637
	0.217	-0.00196	0.163	-0.00541

EXC. RB 2019

ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	176	176	176	176
Average CRPS	2.654		2.645	
Inside 90 % Interval	89.773		89.773	
Inside 95 % Interval	96.59		94.886	
Mean	0.026	-0.01621	0.015	-0.01621
Root-Mean Square	4.916	0.50862	4.908	0.50862
Mean Standardized	0.003	-0.03135	0.0006	-0.03134
Root Mean Sq Standardized	0.947	0.9931	0.944	0.99252
Average Standard Error	5.175	0.51215	5.176	0.51245
	0.259	0.00353	0.268	0.00383

EXC. RB 2020

ERROR	EXPONENTIAL		K-Bessel	
	EBK	Indicator	EBK	Indicator
Count	220	220	220	220
Average CRPS	4.701		4.683	
Inside 90 % Interval	88.182		88.182	
Inside 95 % Interval	95		96.818	
Mean	0.187	0.00501	0.157	0.00587
Root-Mean Square	9.288	0.50127	9.269	0.49835
Mean Standardized	-0.0013	0.00876	-0.0038	0.00984
Root Mean Sq Standardized	0.975	0.98922	0.959	0.99017
Average Standard Error	9.912	0.50664	10.098	0.50255
	0.624	0.00537	0.829	0.0042

Fig. 6.4: Cross-validation result for excavation roadway and borrow-comparison

As shown in the figures above, the cross-validation analysis was started by comparing embankment 2015 values, for EBK exponential and indicator exponential, and the EBK k-Bessel and indicator k-Bessel. As shown in the figure, it is seen that in both classes, when comparing the root mean square to the average standard error, the values for indicator kriging are closer than the values for the EBK. For EBK exponential in this year, the indicator value for root mean square standardized is shown to be closer to one as compared to the EBK, and both the mean values, mean and mean standardized were closer to zero as compared to the EBK which makes the indicator exponential to be chosen for embankment 2015. Similarly, For the k-Bessel semi-variogram, the root mean square of the k-Bessel is closer to one for the indicator than the EBK, and the mean values mean standardized, and mean are closer to zero for the indicator than the EBK. This again makes indicator k-Bessel to be chosen for embankment 2015.

For embankment 2016, within the indicator kriging, the error values for the two models, exponential and k-Bessel seem to show a similar result. A separate comparison was done for the EBK and the indicator kriging. Again, as shown in the figure the difference between root mean square and average standard error was lower for the indicator than the EBK. On the analysis for the exponential for the indicator and EBK, the root mean square standardized for the indicator is close to one than the EBK. And the mean values for the indicator are closer to zero than they are with EBK. This leads to the conclusion of choosing the indicator exponential for embankment 2016. Similarly, for the k-Bessel semi-variogram again the root mean square standardized for the indicator is closer to 1 than the EBK, and the mean values for the indicator are closer to zero than the EBK counterpoints, for this reason, indicator K-Bessel for embankment 2016 was chosen.

For embankment 2017, both for the exponential and k-Bessel, the difference between root mean square and average standard error is lower for the indicator than the EBK. For the exponential, as shown in the figure, the root mean square standardized for the indicator is closer to one than that for the EBK, and the mean and mean standardized for the indicator is closer to one than EBK so indicator exponential for embankment 2017 was chosen. For the k-Bessel even though the root mean square standardized is closer to one for EBK than the indicator since the mean and mean standardized are closer to zero indicator k-Bessel for Embankment 2017 was chosen. By similar approach for embankment 2018, the analysis of the results showed that indicator exponential and indicator k-Bessel were found to be the better models. For embankment 2019, both for the exponential and k-Bessel, the difference between root mean square and average standard error is lower for the indicator than the EBK. For the exponential, since the root mean square standardized is closer to one for indicator than the EBK, and the mean and mean standardized of indicator is closer to zero, indicator exponential for embankment 2019 was chosen. Similarly, for the k-Bessel semi-variogram, the root mean square standardized for the indicator is shown to be closer to one than the EBK, and the mean values for the indicator are closer to zero than the EBK counterpoints. For this reason, the study was led to choose indicator k-Bessel for embankment 2019.

For embankment 2020, both for the exponential and k-Bessel, the difference between root mean square and average standard error is lower for the indicator than the EBK. For the exponential, since the root mean square standardized is closer to one for indicator than the EBK, and the mean and mean standardized of indicator is closer to zero, indicator exponential for embankment 2020 is chosen. Similarly, for the k-Bessel semi-variogram, the root mean square standardized for the indicator is closer to one than the EBK, and the mean values for the indicator are closer to zero than the EBK counterpoints, for this reason, indicator k-Bessel for embankment 2020 is chosen.

A similar approach was followed for excavation boulder and excavation roadway and borrow unit items of work. For excavation boulder, exponential semi-variogram, for all years from 2015-2020, indicator kriging performed better than the EBK with less error, making the indicator exponential to be chosen for excavation boulder 2015- 2020. Similarly, for the k-Bessel, indicator k-Bessel for excavation boulder 2015- 2020 was chosen. For the excavation roadway and borrow unit of work exponential, again following a similar approach, indicator exponential for excavation roadway and borrow 2015- 2020 and indicator k-Bessel for excavation roadway and borrow 2015- 2020 were chosen. The assessment done indicates that for the chosen bid items, the indicator kriging seems to better predict the cost data with less error with both kinds of semi-variograms.

6.4. Conclusion

Accurate conceptual cost estimation of earthwork is a challenging activity. Earthwork is considered as a major construction activity of a highway construction. At the initial stage of construction, because of the limited data available, it is challenging to estimate the conceptual cost of highway earthwork accurately. Accurate estimation of highway earthwork cost is critical because of its high magnitude. The study conducted an estimation method that compared empirical Bayesian kriging with indicator kriging for the same type of semi variograms- exponential and k-Bessel. Three highway earthwork bid items were collected from the year 2015 to 2020 from the Iowa Department of Transportation. The inflation factor was used to bring the price values for each year to 2021. The analysis from the cross-validation represents that the indicator kriging performs better than the empirical Bayesian kriging with less error to predict the cost for the unit items of work presented under the study, for both exponential and k-Bessel semi variograms. This paper contributes to the existing knowledge, the application of indicator kriging and empirical Bayesian kriging, by comparing the prediction errors associated with the exponential and k-Bessel semi-variograms to choose the more accurate early estimation approach with less error. This study can be used as a base for future studies to apply different kriging methods with different semi variogram combinations to bring a better accurate estimation method.

6.5. References

Ali, E. (2020). Geographic Information System (GIS): Definition, Development, Applications & Components. Department of Geography, Ananda Chandra College. India.

Asa, E., Saafi, M., Membah, J., & Billa, A. (2012). Comparison of linear and nonlinear kriging methods for characterization and interpolation of soil data. Journal of Computing in Civil Engineering, 26(1), 11-18.

Askew, W. H., Al-Jibouri, S. H., Mawdesley, M. J., & Patterson, D. E. (2002). Planning linear construction projects: automated method for the generation of earthwork activities. Automation in construction, 11(6), 643-653.

Krivoruchko, K. (2001). Using linear and non-linear kriging interpolators to produce probability maps. Available from ESRI online at http://www. esri. com/software/arcgis/arcgisxtensions/geostatistical/research_papers. html.

Lloyd, C. D., & Atkinson, P. M. (2001). Assessing uncertainty in estimates with ordinary and indicator kriging. Computers & Geosciences, 27(8), 929-937.

Luo, W., Xuan, T., & Li, J. (2017). Estimating anisotropic soil properties using Bayesian Kriging. In Geo-Risk 2017 (pp. 370-381).

Moyeed, R. A., & Papritz, A. (2002). An empirical comparison of kriging methods for nonlinear spatial point prediction. Mathematical Geology, 34(4), 365-386.

Oberlender, G. D., & Trost, S. M. (2001). Predicting accuracy of early cost estimates based on estimate quality. Journal of construction engineering and management, 127(3), 173-182.

Olea, R. A. (2006). A six-step practical approach to semi variogram modeling. Stochastic Environmental Research and Risk Assessment, 20(5), 307-318.

Sodikov, J. (2005). Cost estimation of highway projects in developing countries: artificial neural network approach. Journal of the Eastern Asia Society for Transportation Studies, 6, 1036-1047.

Zitta, N., Aliyu, M. R., Peter, E. U., & Ayuba, Y. (2019). Estimating Cost of Earthwork Using Computerized Visual Display Approach. ATBU Journal of Science, Technology and Education, 7(1), 138-148.

7. CONCLUSION AND RECOMMENDATION

It is challenging to estimate highway earthwork costs at the initial stage of the project. There is only a limited amount of study conducted on the conceptual cost estimation of highway earthwork. The objective of this paper is to predict highway earthwork cost accurately at the conceptual stage of construction by using two spatial statistical methods-empirical Bayesian kriging and indicator kriging methods together with exponential, whittle, and k-Bessel semi-variograms. Iowa Department of Transportation was the source of the earthwork cost data providing unit costs for the three bid items: embankment, excavation boulder, and excavation roadway and borrow for the years 2015 to 2020. The prices were brought to the year 2021 by considering inflation.

The research is divided into seven chapters. In the first chapter of the introduction, the paper introduced the background of highway earthwork construction and conceptual cost estimation. This chapter presented the research questions that were addressed in the subsequent chapters. The second chapter is the systematic literature review chapter which addressed research questions one and two. The findings for research question one suggests that not enough literature was conducted on the conceptual cost estimation of highway earthwork construction projects. But few papers were collected that used computerized systems to estimate the cost of highway earthwork cost, others used regression models, artificial neural networks, and mathematical equations. The common phenomenon that was witnessed in the majority of the studies was the scarcity of data in which a better organization of data was recommended. For research question two, different factors were identified from different studies that affect the conceptual cos estimation of highway earthwork projects. Amongst the factors, the location factor was the one that was seen repeatedly on the articles. This creates the base of this study, finding means of prediction that incorporates the effect of location. The third chapter was exploratory data analysis and, in this chapter, the trend of the cost data from the Iowa Department of Transportation was carefully analyzed. The findings indicate that several outliers which are extreme values out of the range of the data and boarder were screened and removed by using histograms and QQ plot in Geographic Information System (GIS) and Grubbs test in Minitab. It was also noticed that the majority of the data points show a distribution that is skewed to the right so normalization was applied on them. In chapter four, to answer research question three, empirical Bayesian kriging was implemented together with exponential, whittle, and k Bessel semi-variograms to compare and choose the more accurate combination that gives less error. The findings show that the exponential detrended semi-variogram performed better than the other two semi-variograms when the cross-validation error values were analyzed for the majority of bid items. In chapter five indicator kriging was applied to the cost bid data in combination with exponential and k-Bessel semi-variograms to answer research question four. Similarly, by the close analysis of the cross-validation error values, the combination of indicator kriging with exponential semi-variogram showed a more accurate result with less amount of error values than the corresponding k-Bessel semi-variogram for the majority of the bid items for the years 2015-2020. In the sixth chapter, empirical Bayesian kriging was compared to the indicator kriging in combination with exponential and k-Bessel semi-variograms. The finding from careful analysis of error values from cross-validation results suggested that indicator kriging performed better with less error for both combinations of exponential and k-Bessel semi-variograms.

The findings from this study will be beneficial to the construction industry by incorporating kriging methods into the conceptual cost estimation process specifically on highway earthwork projects by considering the effect of location. The comparison of the various kriging methods presented in this study will assist in increasing accuracy in prediction by reducing error. Even though the study assists the conceptual cost estimation, it has limitations. This study used only three available bid items. So future studies can perform the same method on various kinds of bid items to cross-check the results. And also, this research is specified only to three semi-variograms; exponential, whittle, and k-Bessel, so further research can be conducted on other variety of semi-variograms.