A Study of Ranking Routes for Electrical Transmission Lines using Weighted Sum Model and Fuzzy Inference System

by

Imtiaz Ahmed Khan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2020

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Selecting a route is the first step in building a new electrical transmission line. The most common practice in selecting a route involves ranking possible route options, which is a complex process demanding many decision considerations be taken into account. Even to this day, the ranking process is mainly done manually by humans using printed maps and field surveys, making it time-consuming and prone to errors.

This thesis studies the most common decision considerations that affect the process of ranking a set of route options and classifies them into four main categories. Then, the work proposes a methodology to automate the process of ranking routes for an electrical transmission line and implements it using Geographic Information System (GIS), image processing techniques, and Weighted Sum Model (WSM). It evaluates the effectiveness of the methodology by comparing the results obtained with industrial results of an actual project in Saskatoon, Canada. The preliminary results are very promising. Out of five route options, the proposed methodology ranks the top two options accurately, and it successfully identifies the least-preferred route options.

To validate the methodology further, the thesis generates synthetic data and tests it with various simulated scenarios. The work generates random routes and hypothetical features, using perturbation and image processing techniques, to test the methodology with more route options and decision considerations, respectively. In the process of validation, it also improves the accuracy of the methodology by refining the WSM. The methodology with refined WSM successfully outputs expected results when tested with one hundred and fifty five routes and taking six decision considerations into account.

The thesis also implements and tests the methodology using Fuzzy Inference System (FIS), instead of WSM, to mimic the human decision process and to allow users to dictate the importance of different decision considerations using linguistic variables. It then compares the two methods and discusses their advantages and disadvantages. Both methods perform well and have around 80% similarity in the outputs produced. However, they are both unique in their own ways. FIS allows users to describe their preferences using linguistic variables making it more user friendly, while, WSM is more predictable and easier to fine tune results. Thus, the thesis presents two ways of automating the process of ranking routes for an electrical transmission line.

Acknowledgements

All praise to almighty Allah, the most powerful and most gracious, for bestowing me with the knowledge and capability to accomplish this thesis. Everything that I am and have achieved is because of him.

I would like to thank my supervisor, Prof. Kshirasagar Naik, who has been a fantastic mentor and a constant source of support and encouragement. His door was always open for me no matter how often I wanted to meet him. I could not have completed this work without his guidance, patience, and valuable feedback. I would also like to extend my sincere thanks to Prof. Mohamed Ahmed for his valuable time and insights. He helped me get access to industrial data and resources. This would not have been possible without our fruitful discussions.

I am grateful to all my friends for their prayers and moral support, especially Muntaka, Ammar, Karthik, and Mohmoud. I am very lucky to have such genuine people in my life. I thank my colleagues, Mustafa Al-tekreeti, Abdurhman Albasir, and Md. Rubayatur Bhuyian for their constant help and guidance.

Last but surely not the least, I would like to thank my mom, dad, my elder brother, and sister-in-law for their continuous support, patience, and prayers. I owe a great deal of gratitude to my wife, Tasnim Mostafa, for bearing with me during my lows, believing in me, and motivating me during the years of my study.

Dedication

This is dedicated to Abbu, my loving father, who did not live long enough to see me graduate.

Table of Contents

List of Tables

List of Figures

Abbreviations

AHP Analytic Hierarchy Process 13

DEM Digital Elevation Model 14

FIS Fuzzy Inference System iii, xiii, 5, 7, 13, 15, 16, 19, 70, 71, 73, 75–77, 81, 82, 87, 91

GA Genetic Algorithm 13

GIS Geographic Information System iii, 5, 7, 12–14, 23, 70, 90

GUI Graphical User Interface 14

MCDM Multi-Criteria Decision Making 13, 15

QGIS Quantum Geographic Information System 14, 23, 30

ROW Rights of Way 11, 34, 75, 76, 78

WSM Weighted Sum Model iii, xiii, 5, 7, 13, 15, 26, 33, 55, 57, 58, 70, 71, 73, 87, 90, 91

WSMr Weighted Sum Model refined 57, 58, 60, 61, 63, 66, 67, 90

Chapter 1

Introduction

We need electricity basically for everything we do nowadays. However, electric power is not generated everywhere [27],[39]. Most common electric power generating stations, such as hydro, nuclear, fossil fuel, and wind, are situated hundreds of miles away from their load centers, for resource availability, safety, and aesthetic reasons [27]. These distances give rise to power transmission systems that consist of different types of electrical transmission lines.

An electrical transmission line is used to transfer electrical energy from one location to another. Due to the massive increase in demand for electricity, a need arises for new transmission lines very frequently. However, building a new transmission line is a multipart process involving route selection, cost-benefit analysis, and optimization of power flow [29]. Route, also known as corridor or path, selection is the first stage of the process of planning new transmission lines, in which designers decide the best areas to cross. This task is not easy as designers need to consider different types of factors when a new transmission line is built [23],[35],[37], [29].

Taking all the decision considerations into account makes the manual route selection process tedious; thus automating the process has become an active research topic. However, most works have focused on finding the least-cost path or optimal path [23],[35],[43],[41]. The problem with this approach is that, in most cases, the optimal path is not feasible or practical. Consequently, decision makers are left without any alternatives. As a result, these methods are not being used in the industry. In practice, to select an optimal path, designers and engineers need to rank a set of feasible route options based on a set of decision considerations, and then choose the most viable one from the top-ranked options [27]. To date, however, little work has been done to automate the process of ranking routes for a given set of decision considerations.

1.1 Motivation

The subject of this thesis was motivated by the following facts:

- *Massive increase in electricity demand:* With modernization, the demand for electricity continues to increase [3]. Figure 1.1 illustrates historical global primary energy consumption in chronological order. From the graph, we can see that the demand for electricity globally is increasing almost exponentially. Factors such as population growth and greater use of electrical appliances and equipment are expected to continue to drive electricity demand in the coming years [19]. Thus, the need for more generating stations and transmission lines in the near future is inevitable.

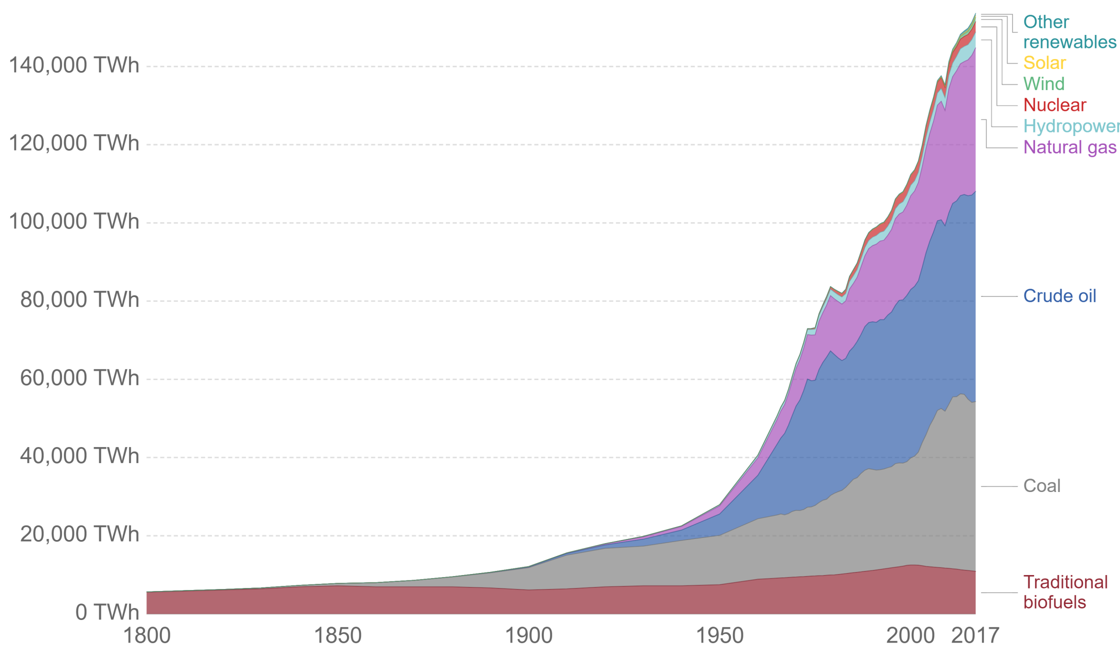


Figure 1.1: Global primary energy consumption, measured in terawatt-hours (TWh) per year. Here 'other renewables' are renewable technologies not including solar, wind, hydropower and traditional bio-fuels [3].

- *Change in the electricity market:* The structure of the electricity sector has been evolving over the past decade [22]. In the past, the electric utilities, which were vertically-integrated, provided electricity with monopoly rights [12]. All electricity producers or utility providers build their own transmission and distribution systems to supply electricity to their customers [11]. They had complete control over generation, transmission, and distribution of electricity along with the price structure for industrial and residential users in their regions. However, this setup was costly because it often led to underused infrastructure and wastage of electricity. Over the past decade, the structure of the electricity industry has undergone significant change. Most provinces have un-bundled the generation, transmission, and distribution functions of electric utilities into separate organizations [22]. Also, some provinces have moved towards a more competitive generation system with the private sector playing an increasing role, giving rise to independent power producers [19]. As a result, the competition of building new transmission lines by the private sector to gain access to the electricity market is also increasing.

- *Advent of smart connected grids:* If we look at Figure 1.1, we can see that over the years, not only did the demand for electricity increase, but the types of source tapped to meet the demands have also increased. In the 1800s, most of the demand for electricity was met with traditional bio-fuel. Nowadays, we have coal, crude oil, natural gas, hydro-power, nuclear, wind, solar, and other renewable sources pumping electrical energy into the system [4]. This called for a connected grid system to utilize existing utility infrastructure and to increase the reliability of electricity supply to consumers [10]. As a consequence, the need for new transmission lines between substations is growing.

- *Complexity in selecting routes:* Selecting a path for a new electrical transmission line is not easy. There are many stakeholders, such as investors, governments, utility providers, environmentalists, indigenous people, engineers and consultants, for every new transmission line [40]. Finding a route that is agreeable to all the parties is challenging. The traditional-manual method of finding the route that cumulatively minimizes the dissatisfaction of all parties requires careful evaluation of every possible route. To do so requires many engineers and consultants specialized in different aspects, which makes the manual method labor-intensive [35]. Moreover, the evaluation needs to be repeated for all other possible routes, making it an iterative and time-consuming process. The manual method of selecting a route can also be erroneous due to inconsistency in human evaluations.

It is evident from these facts that the need to build new transmission lines in the near future is inevitable. It is also clear that automating the route selection will expedite the whole process. Thus, this thesis is motivated to study the process of selecting routes for electrical transmission lines and propose a methodology to automate it.

1.2 Problem Statement

Ranking route options for an electrical transmission line is complex. Manually ranking them is iterative, time-consuming and labour intensive [35]. It requires large amounts of detailed information in the form of maps and field surveys to be collected, analyzed, and evaluated for every possible route option by various engineers and consultants. Moreover, different maps may use different projection systems [21], [31], [1], [18]. Thus, the manual method is prone to errors.

Automating the process of ranking transmission line options will make the selection process faster and more standardized. This will allow more route options to be evaluated and increase the chance of selecting the best feasible path. It will also reduce project revision efforts [35]. Thus, the problem this thesis is trying to solve is how to automate the route selection process for electrical transmission lines and what decision considerations we need to take into account.

1.3 Solution Strategy and Contributions

To address the problem and propose a solution, we need to formally articulate the problem. For our purpose, a geographic area is an area of earth with natural and man-made features. Examples of natural features are lakes, rivers, mountains, reserves, and forests. Examples of man-made features are roads, railways, residential areas, commercial areas, buildings, private land, historic landmarks, and parks. A map is used to represent a geographic area and its features on a 2-dimensional plane [17]. Different types of maps represent different types of features. For example, a political map displays the man-made boundaries, such as state and national borders. A physical map displays the physical features, such as mountains, plains, rivers and oceans, whereas, an economic or a resource map displays the arrangement of natural resources and the economic activity within a place. A road map shows roads and highways in varying levels of detail. It also indicates important natural and man-made features, such as cities and national parks.

Today, all these types of maps are available in digital formats, namely vector and raster formats (discussed in more detail in Chapter 2, section 2.5.2). The advantage of digital maps is that a combination of different types of maps can be superimposed to render a new map with only features of interest [32], [7]. Figure 1.2 shows an example of a digital map with features such as roads, railways, rivers, parks, residential areas, and commercial areas. The geographic area of interest is the minimum rectangular area that encapsulates the geographic region we need to consider.

In Figure 1.2, points B and E represent a generating sub-station and a load center, respectively. An electrical transmission line is used to transfer electrical energy from one location to another. The path it takes from the generating station to the load center is called the transmission line route, path, or corridor, interchangeably. The corridor needs to have enough vertical and horizontal clearance for the transmission line to meet engineering and regional standards [27]. Possible route options can be represented on the map using colored lines, as shown in Figure 1.2.

Each feature, natural or man-made, along the path of the transmission line introduces levels of complexity that differ according to the circumstances while constructing or maintaining the electrical transmission line. For example, the presence of a mountain or private land in the path may increase costs as it may demand additional supporting structures or compensation for the land use. On the other hand, a road or railway coexisting along the transmission line corridor may reduce the complexity of construction and maintenance by providing easy access. As a result, multiple paths need to be assessed based on these decision considerations before selecting an optimal path, to minimize engineering complexity and economic cost while mitigating adverse environmental and social impacts [27], [8].

Thus, our objective is to rank a set of feasible route options based on a given set of decision considerations and their weights of importance. Here, the weights define the importance of a decision consideration compared to others for decision makers. Ranking can be achieved with a four step methodology, proposed and described in detail in Chapter 3. The thesis implements and tests the methodology using GIS, image processing techniques, and WSM. It also implements the methodology using FIS, instead of WSM, to mimic human knowledge and approximate reasoning in the decision process. It then compares the two methods.

In summary, this research makes the following contributions:

- It proposes a methodology for ranking a list of route options for an electrical transmission line;
- The methodology has been implemented using WSM and validated;
- It has also been implemented using FIS and validated; and

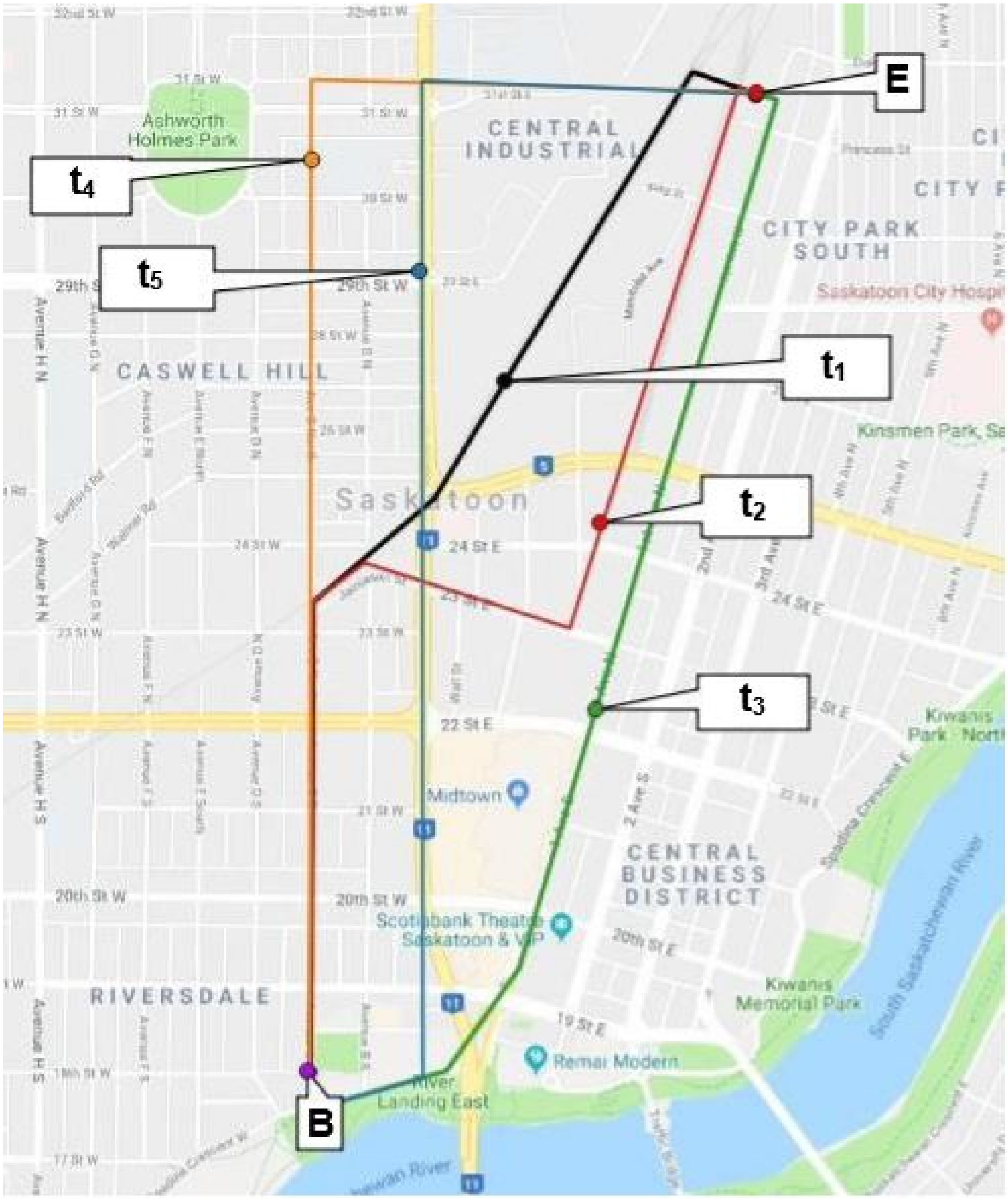


Figure 1.2: The map M that shows five feasible route options for a transmission line to connect the two regions B and E.

- The two methods have been compared.

1.4 Thesis Organization

The remainder of the thesis is organized as follows: Chapter 2 presents a brief background on topics such as the electricity market in North America, decision considerations used by engineers and consultants in ranking transmission line routes, and tools and techniques used in the thesis that might help in understanding the work better. It also discusses related works by other researchers. Chapter 3 introduces the proposed methodology and explains all the steps in detail. It implements and tests the methodology using GIS, image processing techniques, and WSM. In Chapter 4, synthetic data is generated to validate the proposed methodology further. Chapter 5 verifies the methodology for different scenarios and analyzes the results. FIS is implemented in the methodology instead of WSM in Chapter 6 and the two methods are compared. Chapter 7 concludes the thesis.

Chapter 2

Background and Related Works

This chapter provides a general introduction to related topics, briefly discusses important tools and techniques used in the research, and reviews existing work by other researchers on the subject. The presentation here is intended to clarify the research work for readers and not be exhaustive.

2.1 Electricity industry in Canada

Electricity is one of the most versatile types of energy that can be easily converted to other types such as thermal, light, chemical, mechanical, and sound energy [13]. Due to its versatility, the demand and need for it in different sectors [2] (industrial, commercial, residential, and transportation) is rising. To meet the demands of electricity and to supply it efficiently at lower cost, electrical power systems were developed [12]. Thus, the electricity industry is involved in three main activities [19]:

- the generation of electricity using various energy sources and technologies;
- the transmission of electricity from power plants to load centers; and,
- the distribution of electricity to end-users at the load centers.

In Canada, the provincial governments are primarily responsible for the generation, transmission and distribution of electricity [19]. They exercise their legal power through Crown utilities and regulatory agencies. In the past, electricity has been provided mainly by vertically-integrated [16] electric utilities, which were often provincial Crown corporations, with monopoly rights. Some large industrial electricity consumers (such as aluminum manufacturers) have also built their own electricity generation facilities to meet their power requirements.

Over the past decade, the structure of the electricity industry has undergone significant change [19], [11]. Most provinces have un-bundled the generation, transmission and distribution functions of electric utilities into separate organizations. Some provinces have also moved towards a more competitive generation system with the private sector playing an increasing role, giving rise to independent power producers. Current drivers of change within the electricity industry in Canada include the increasing use of natural gas to generate electric power; distributed generation with interconnected grids; release of FERC 1000 order; growing implementation of renewable energy and the retirement of coal and nuclear generation; and increasing interactions at the Federal, state, and local levels [11]. Internal drivers of change rose mainly from the development and implementation of smart technologies, which are increasing the ability to maximize the use of connected grid resources, improving grid productivity to control power flows, troubleshoot problems remotely, and allow customers to better manage their energy use [11]. The structural change of the electricity market is increasing investments in the large transmission and distribution sectors to replace the aging infrastructure with more reliable and market efficient interconnected network.

2.2 Route selection process

Route selection is one of the first step in building a new transmission line. Figure 2.1 shows a typical transmission route (corridor) selection procedure used in the industry. The general route selection objectives are: minimize adverse effects to sensitive environmental resources; minimize adverse effects to significant cultural resources (archaeological and historical); minimize adverse effects on designated scenic resources; minimize conflicts with local, state, and federal land use plans and resource policies; minimize the need to acquire property; maintain public health and safety; maximize the reasonable, practical, and feasible use of existing linear corridors (e.g., transmission line, highways, railways, pipelines); comply with all statutory requirements, regulations, and state and federal siting agency policies; and achieve a reliable, operable, and cost-effective solution [8]. These objectives form the basis in the considerations taken into account while selecting a route. Today, the planner selects the appropriate transmission route based on his or her knowledge of the system, the results of the system analysis, and available rights of way [27].

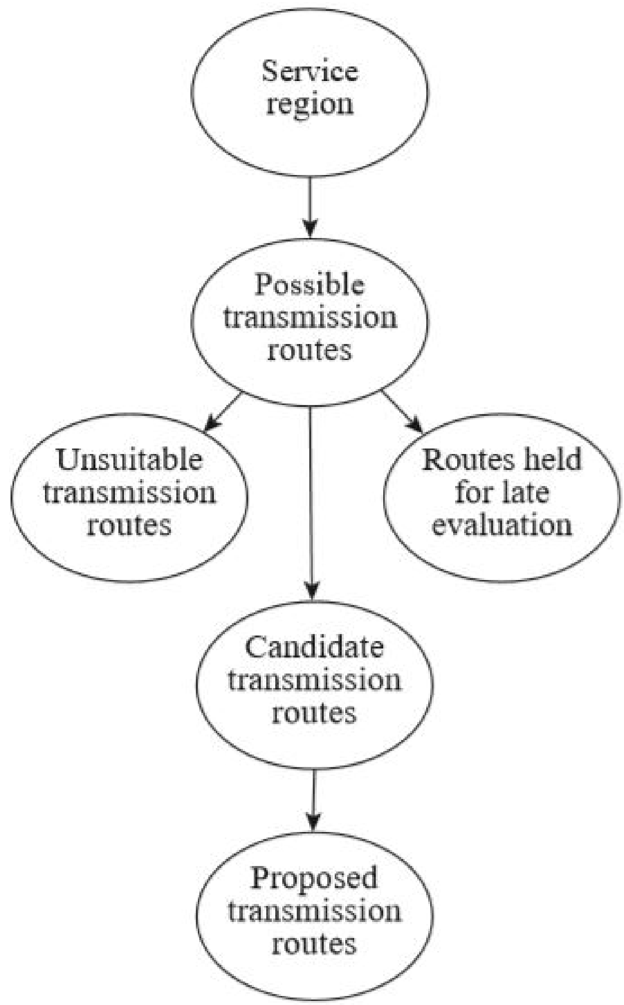


Figure 2.1: Transmission route selection procedure [27].

2.3 Decision considerations

This section explores the main factors that are taken into consideration while ranking available route options. The considerations can be broadly classified into four categories, engineering, economic, environmental, and social, based on their impact. For instance, if a decision consideration has an impact on cost, the thesis categorizes it as an economical consideration, while if it has an effect on the environment, the thesis categorizes it as an environmental consideration.

2.3.1 Engineering considerations

Engineering considerations simplify construction, operation, and maintenance of a transmission line from the technical perspective. For example, the accessibility of the line path via a road is an important factor in determining which path to select, because without a suitable road, it will be difficult to transport large structures to the site, such as steel towers, transformers, and other electrical equipment [23]. Similarly, having many sharp turns in the path will increase the need for stronger towers and supporting structures [35].

2.3.2 Economic considerations

Economic considerations try to minimize the overall cost of a project. For example, the length of a path is an important economic consideration because the number of needed reels of conducting wires, steel towers, and structures increases with length [27]. It also increases the cost of maintaining the transmission line. Another important decision consideration is the use of an existing Rights of Way (ROW) [27]. The ROW is the legal permission given to an entity to access private or state land and use it to construct and maintain a public utility. Using existing ROW from other facilities, such as roads, railways, telephone lines, or pipelines minimizes the need to compensate for land use and saves time in getting approval [27].

2.3.3 Environmental considerations

Environmental considerations try to minimize a project's effect on the environment. For example, a transmission line going through a sensitive ecosystem, such as forests or wetlands, can adversely affect the flora and fauna in it [37]. As a result, an alternative route should be considered, bypassing the sensitive area. This is also true for historical landmarks and protected areas. Most transmission line towers are very tall and can disrupt the visual aesthetics of an area [27]. Hence, these factors need to be addressed while deciding on the path.

2.3.4 Social considerations

Social considerations take the policies and opinions of political, legislative, and indigenous groups into account [35],[37]. For example, if a path crosses through private land or restricted land that is of importance to the indigenous people, it will be opposed by that community [37]. Thus,intrusion into such areas needs to be excluded or at least minimized along the path. Moreover, there can be laws and restrictions from different governing bodies on whether a transmission line can pass over areas under their jurisdiction or not [35]. Such factors need to be taken into consideration while selecting the optimal path.

2.4 Related Works

In the quest for automating the route selection process for electrical transmission lines, many researchers have suggested different decision considerations and processing techniques. In reference [43], the authors took into account forest, agriculture, flora, wildlife, hydrology, landscape, and mining resources, while in reference [37], they considered housing density in the area of interest, state regulations, low elevation area, distance to the road, and impacts on the natural environment. The authors of [35] have considered implementation and maintenance costs associated with slope, soil, vegetation, and land ownership. They also tried to minimize the number of bends in the resulting paths. In reference [20], distances to routes from buildings, hospitals, and schools, and visibility from cultural and recreational sites were considered. On the other hand, some researches, such as [33] and [36]have considered weather data to avoid lightning-prone areas and temperature hot-spots, respectively.

Numerous ingenious methods and techniques have been proposed in the literature to automate the process of selecting a suitable route for an electrical transmission line. For example, the authors of [43] proposed a method to find the optimal transmission route using image processing techniques on satellite images that identified areas with environmental constraints. The authors of references [41], [44], and [34] showed the potential of GISs in solving engineering design problems. They used GIS to automate the design of an electrical supply system in rural areas and the planning of electrical distribution networks using terrain information. These efforts paved the way for many other important works in which GIS was used to find the optimal path for a transmission line. The authors of [23] and [37] used GIS along with Analytic Hierarchy Process (AHP) – a type of Multi-Criteria Decision Making (MCDM) – to find the least cost path. In reference [35], the authors used raster data, rather than vector data in GIS, which allowed them to adopt dynamic programming to find the optimized route. They proposed a method that associates constraints with installation, operation, and maintenance cost. Many works, such as [20], have proposed the use of least-cost path analysis to find optimal paths.

In addition, other works have proposed the use of advanced techniques, such as machine learning, genetic algorithms, and bee colony algorithm to find the optimized path for an electrical transmission line. In reference [24], the authors used Q-learning, a type of machine learning algorithm, to find the optimal route. In [25], the authors used Genetic Algorithm (GA) on raster-based maps to find the best route. A recent work [26] has proposed improved genetic and artificial bee colony algorithms. Most of these works have tried to either find the least-cost path or the optimal path. However, little work has been done on ranking a set of feasible paths. Although the authors in reference [38] have mentioned finding multiple alternative paths, they mainly focused on finding non-overlapping routes for building more than one transmission line simultaneously.

2.5 Tools and techniques

This section briefly discusses the techniques and tools used in this thesis. In this work, the main objective is to automate the process of ranking a given set of route options for an electrical transmission line, based on a set of decision considerations. To achieve this goal, the research utilizes the GIS, image processing techniques, and the WSM, described in detail in Chapter 3. Later in Chapter 6, it implements FIS, instead of WSM, to test an alternative way of fulfilling its goal. These tools and techniques will be discussed briefly in the following sub-sections.

2.5.1 Geographic Information System

The GIS is used for capturing, storing, checking, manipulating, analyzing, and displaying data that are spatially referenced on earth [23],[35],[37]. It has been used throughout the world in global, regional, and local environmental studies [24], and it has the potential to be used in applied science and engineering for accurate geospatial data representation and computation.

In GIS, spatially referenced data are mainly represented by two formats: vector and raster [15]. Vector format data use points, lines, polygons to represent features. For example, a path can be represented with a line and a forest area can be represented with a polygon on a map. The common types of vector files are shapefile (.shp or .shx or .dbf), geodatabase (.gdb), Keyhole Markup Language (KML/KMZ), and OpenStreamMap (.osm)[9]. Alternatively, raster format is a type of image that uses pixels to represent a feature or information about a feature. For example, a Digital Elevation Model (DEM), which is a type of raster map, stores an altitude value in each pixel representing a subdivision of a given area. The common types of raster files are Joint Photographic Experts Group (.jpeg), Tagged Image Format File (.tiff), and Graphics Interchange Format (.GIF) [14].

This work uses the open source QGIS desktop software package version 3.2.3. Quantum Geographic Information System (QGIS) is a free and open-source cross-platform desktop application that supports viewing, editing, and analysis of geospatial data [7]. Figure 2.2 demonstrates the Graphical User Interface (GUI) of QGIS, where box 1 shows the layer list and box 2 shows the browser panel. These panels are used to browse and select different layers, representing various features, to display them on the map canvas (box 6). Many commonly used functions are found in the toolbars shown in box 3. QGIS also has many build-in GIS data processing tools that are easily accessible from box 4. Box 5 shows the status bar that displays information that are useful while working with different types of GIS data.

2.5.2 Image processing

To analyze GIS data, the work utilizes simple image processing techniques. It compares binary images using logical operations, and it utilizes a thinning algorithm to skeletonize a line that represents a route option as a one-pixel-wide line. In general, a digital image or a picture is made up of pixels. Each pixel has a value, or values, depending on the type of image [28], such as RGB (Red, Green, Blue), grayscale, raster, and binary. An RGB image is a color image with each pixel having three values that range from 0 to 255. Every value represents the intensity of the color. Similarly, a grayscale image is a black and white image with every pixel value ranging from 0 to 255. Raster images are a special type of images where pixel values can be any real value [23], whereas, pixels in binary images can only be 0s or 1s. Generally, images are stored as 2-D arrays of pixels, which makes image processing tasks more convenient using standard linear algebra techniques [28].

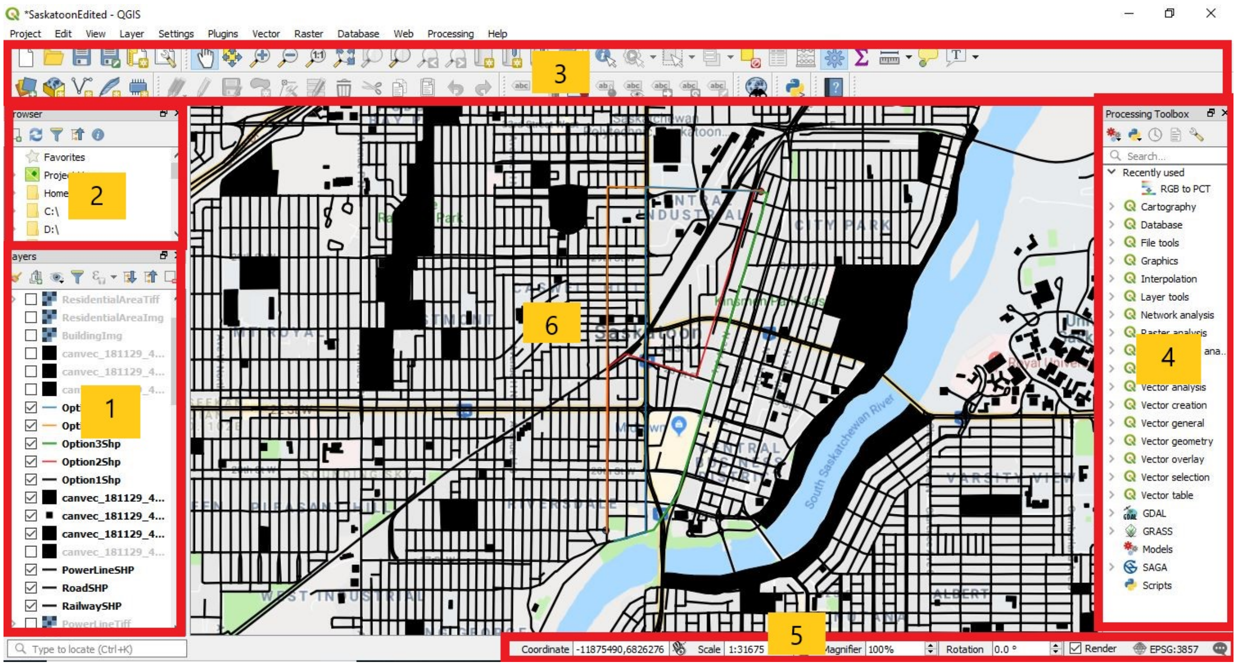


Figure 2.2: A snapshot of the GUI of QGIS

2.5.3 Weighted Sum Model

To score and rank route options, this thesis uses the WSM. This method is particularly powerful when we have a number of good alternatives to choose from and many different factors to take into account[42]. This method, a type of MCDM method, is based on the additive utility assumption [42]. That is, the alternative that has the highest total score after summing up all factor values is the best option and vice-versa. It is a common technique used by our industry partner to rank route alternatives.

2.5.4 Fuzzy Inference System

As an alternative to WSM, to score route options, this research uses a FIS. The FIS tries to mimic human knowledge and approximate reasoning in the decision process. Since the work wants to automate the ranking process that replicates the current manual method, FIS provides an excellent platform for the research goals.

Figure 2.3 shows the block diagram of a FIS where the Rule Base contains fuzzy If-Then rules and the Database defines the membership functions of fuzzy sets used in the fuzzy

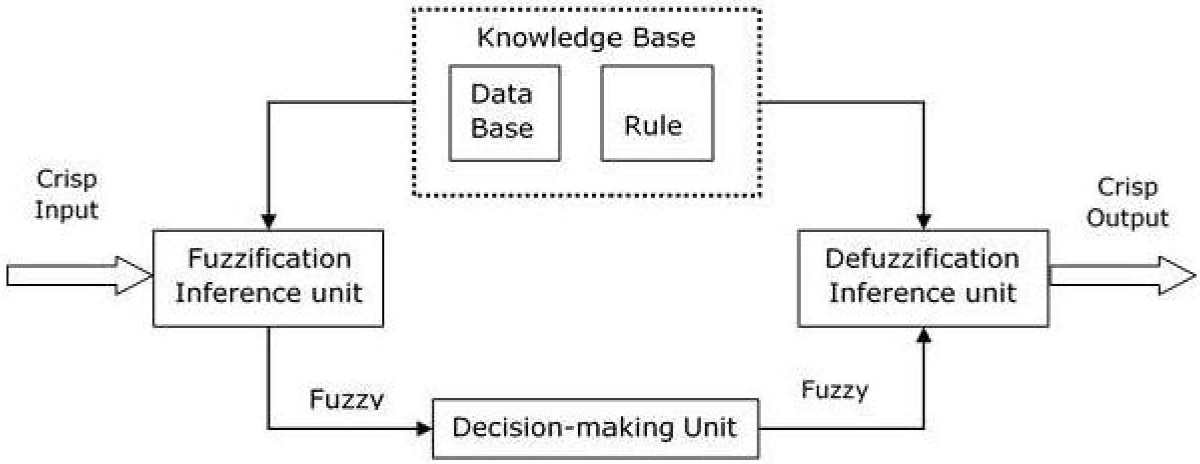


Figure 2.3: Functional blocks of FIS [5].

rules. These two blocks form the Knowledge base of the system. The Fuzzification Inference Unit converts the crisp input quantities into fuzzy quantities and the Defuzzification Inference Unit converts the fuzzy quantities into crisp output. The Decision-making unit performs the operation on rules [5]. A FIS can use different methods to make inferences. The most common methods among them are Mamdani Fuzzy Inference Model and Sugeno Fuzzy Model. These concepts are discussed in more detail below.

Fuzzy sets

The traditional crisp sets consist of crisp variables that represent precise quantities, for example, $x = 3.1415296$, temp $= 55.0°$ F, and $A \in \{0, 1\}$. However, humans use linguistic variables, such as temp: {freezing, cool, warm, hot} and speed: {slow, fast} instead of mentioning the exact temperature and speed in their daily life. Thus, these linguistic variables cannot be represented properly using crisp sets. In contrast, fuzzy sets can take these imprecision in linguistic variables into account. They represent the degree to which a quality is possessed. For example, a temperature $T°$ F can either belong to a crisp set "Warm Temp" or not belong to it. However, in fuzzy logic, the temperature $T°$ F can have a degree to which it belongs to a fuzzy set. The degree of belongingness are represented using fuzzy membership functions.

Fuzzy membership functions

In fuzzy sets, the fuzzy variables have values in the range of [0,1]. These values represent the degree of truth or "membership". Figure 2.4 illustrates an example of membership functions of fuzzy variables for temperature. We can see from the diagram that a crisp temperature of 37° F belongs to two fuzzy variables, "Freezing" and "Cool", with a membership value of 0.7 and 0.3, respectively. The membership functions can be of any shape, but the most common are triangular, trapezoidal, Gaussian, and generalized bell membership functions shown in Figure 2.5.

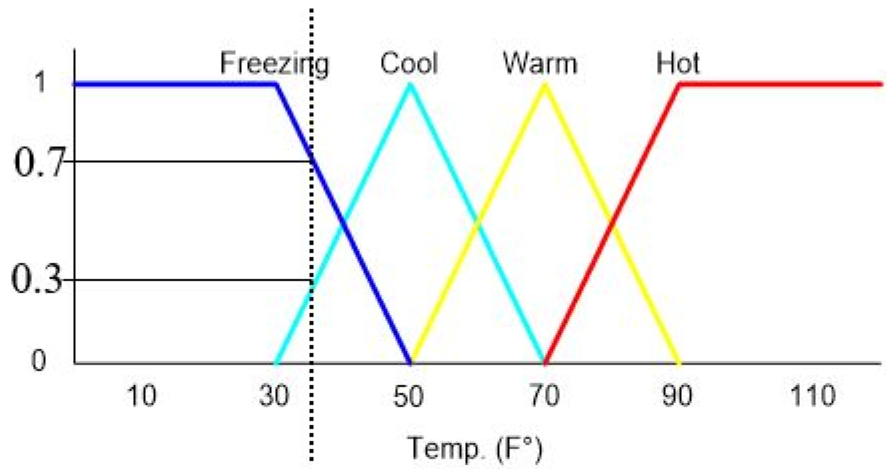


Figure 2.4: An example fuzzy membership functions [5].

Fuzzy logic

Fuzzy logic is a way to make use of natural language in logic. It allows for a realistic extension of binary, crisp logic to qualitative, subjective, and approximate situations [30]. The conventional binary logic is crisp and allows for only two states, true (1) or false (0). This logic cannot handle fuzzy variables, examples of which are "true", "very true", and "somewhat true". Such variables are represented using fuzzy membership functions. In fuzzy logic, the knowledge base is represented by If-Then rules [30], where the fuzzy variables are connected using fuzzy OR (conjunction) and fuzzy AND (disjunctive).

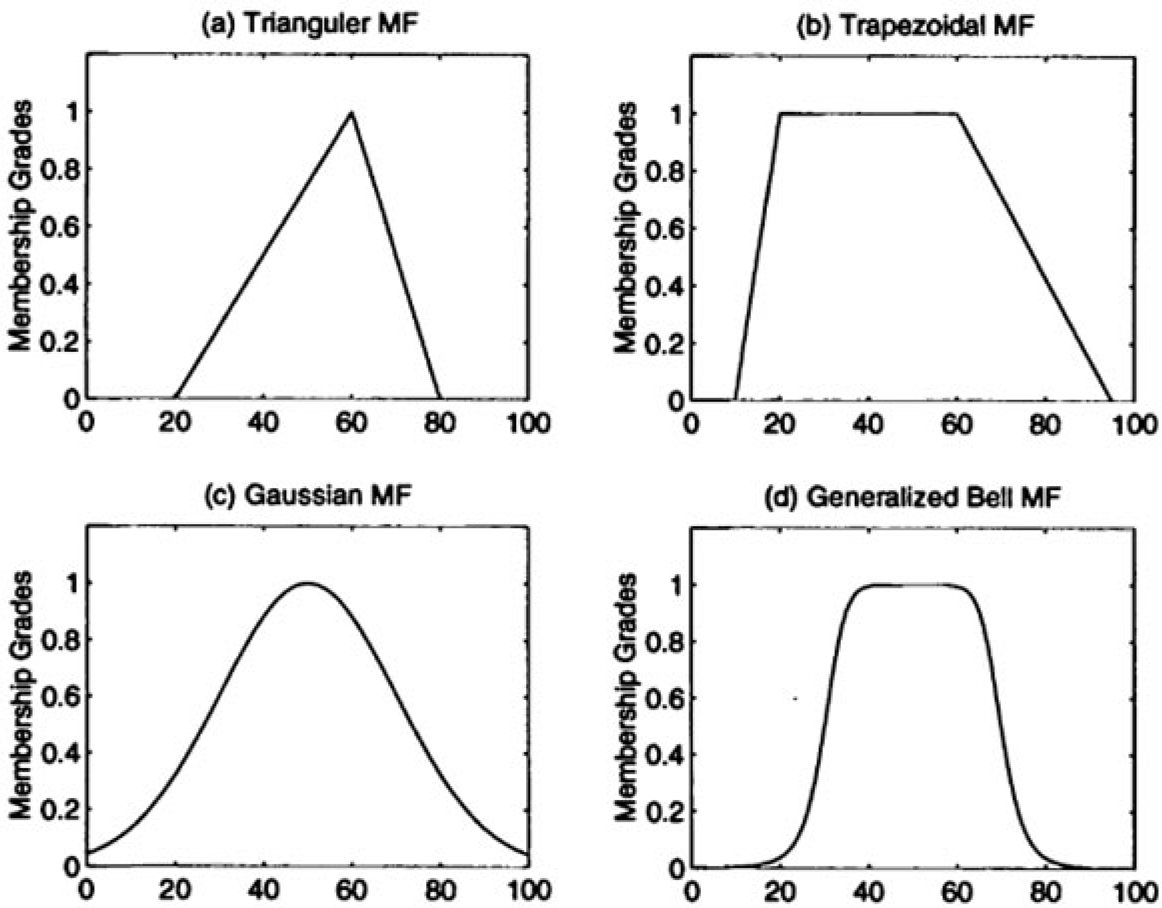


Figure 2.5: Common types of membership functions [6].

Inference model

To infer the fuzzy connectives, fuzzify inputs and defuzzify outputs, the FIS uses inference model. Different inferencing procedures have been used in the literature. However, the most popular are the Mamdani and Sugeno inference models. The main difference among them is the aggregation and defuzzification steps. This work uses the Mamdani model. These concepts and techniques will be frequently referred to in this thesis.

Chapter 3

The Proposed Methodology

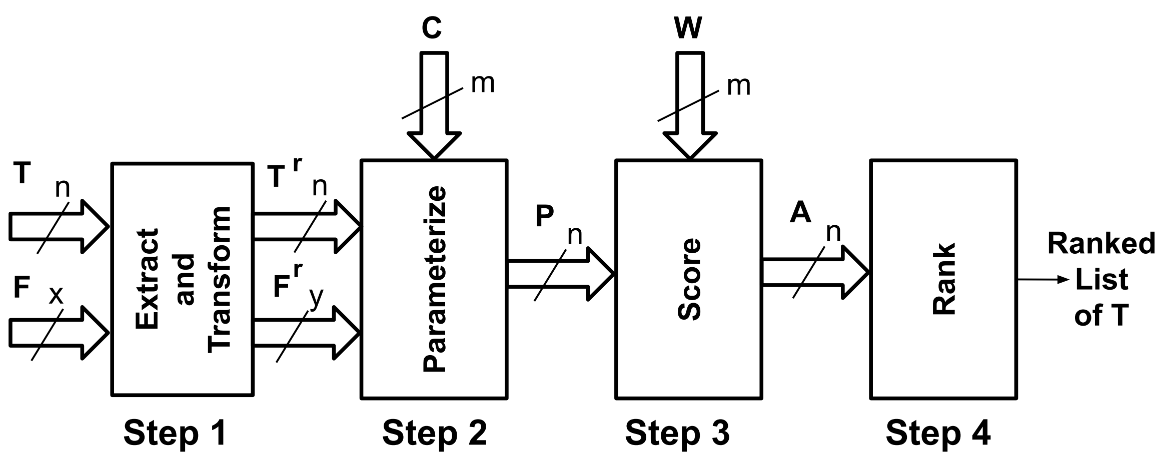


Figure 3.1: The proposed methodology.

Figure 3.1 shows the main steps, the inputs, and the expected output of the proposed methodology. Table 3.1 lists the symbols used in this work. As shown in Figure 3.1, the methodology consists of four different steps. The methodology's output is a list of route options ranked in descending order. First, section 3.1 defines the input requirements. Then, the methodology steps are explained in detail in section 3.2. Section 3.3 demonstrates the methodology using a real example, and finally discusses the outcomes in section 3.4 .

Table 3.1: Nomenclature for proposed methodology

Symbol	**Description**
F	a set of all natural and man-made features
x	number of features in F, $x = \|F\|$
F^e	a set of extracted features
y	number of features in F^e, $y = \|F^e\|$
f_k	an extracted feature belonging to F^e, $1 \leq k \leq y$
T	set of all route options
n	number of route options in T, $n = \|T\|$
t_i	a route option in the set T, $1 \leq i \leq n$
T^r	a set of transformed route options in raster format
t_i^r	a route option representing t_i in raster format
F^r	a set of extracted and transformed features in raster format
f_k^r	an extracted and transformed feature representing f_k in raster format
C	set of all decision considerations
m	number of decision considerations in set C, $m = \|C\|$
c_j	a decision consideration in the set C, $1 \leq j \leq m$
P	a set of parameters describing characteristics of route options
p_{ij}	a parameter describing a characteristic of t_i for decision consideration c_j
W	weights of importance for decision considerations
w_j	weight of importance for decision consideration c_j
S	set of individual scores of all t_i for every c_j
s_{ij}	a score of t_i for c_j
A	set of all total scores of route options
a_i	total score for route option t_i
L	actual length on earth for each pixel
N	maximum score a route option t_i can get for a consideration c_j

3.1 Inputs to the methodology

Let the map M be a digital map of a geographic area of interest with data representing all the natural and man-made features. Let the set of features (natural and man-made) be $F = \{f_k\}_{k=1}^{x}$, where x represents the number of features. Let the set of feasible route options between points B and E be denoted by T. Each route option $t_i \in T$ denotes a possible route for the new transmission line. For example, in Figure 1.2, the set T consists of five possible paths, i.e., $T = \{t_1, t_2, t_3, t_4, t_5\}$. To select the optimal path, we need to rank the elements in T.

Let all types of decision considerations be represented by the set C. Each decision consideration $c_j \in C$ denotes a constraint or opportunity. The decision considerations in C vary from project to project, and are set by the decision makers. For example, in Figure 1.2, we can see that there are no forests or historical landmarks in the geographic area represented by the map M. Hence, the designers may not need to consider these environmental or social considerations for a transmission line in that region. Each c_j has a weight w_j associated with it, where $w_j \in W$, the set of weights that are selected by the decision makers based on the importance of a decision consideration with respect to other decision considerations in C. The importance of a decision consideration is directly proportional to its weight. A larger weight implies high importance of that decision consideration. The work assumes that the feasible route options in T are given in the vector format. It also assumes that the geospatial data in F are in vector or raster formats.

3.2 Methodology procedures

3.2.1 Feature extraction and transformation

Real geospatial data come mainly in two formats (vector and raster) and have much redundant data unimportant for our objective. Thus, in this step, the methodology cleans the data by extracting only the necessary features, and converts the data to a common format by transforming them to the raster format. We want all the data in raster format as image processing techniques cannot be applied to maps in vector format. Formally, this step can be summarized as follows:

$$F^e = \xi(F),\ F^e \subseteq F, \tag{3.1}$$

$$F^r = \{f_k^r : f_k^r = \eta(f_k),\ f_k \in F^e, 1 \leq k \leq |F^e|\}, \tag{3.2}$$

$$T^r = \{t_i^r : t_i^r = \eta(t_i),\ t_i \in T, 1 \leq i \leq |T|\} \tag{3.3}$$

where F^e is the set of extracted features, $\xi()$ is the extraction procedure, $\eta()$ is the transformation procedure, and $|X|$ is the cardinality of the set X. The sets F^e and T in raster format are represented by F^r and T^r, respectively.

To extract features of interest, we carefully *choose* a subset of features from F and *save* it as the set F^e. The extraction process can be attained using a GIS by selecting the features of interest from the the "Layer" toolbox and saving them in a separate folder. An example of the extraction process using "Layer" toolbox in QGIS is illustrated in Figure 3.2. After feature extraction, it is important to transform all the data, T and F^e, to raster files. Raster files are image files with pixels and can be represented as a 2-dimensional matrix of size $U \times V$ pixels, where U represents the number of rows and V represents the number of columns in the image. Since the research is only interested in the presence or absence of a feature with respect to a path in T, raster files are expressed as binary files. That means, the pixel value in the raster file is either 1, representing the presence of the feature, or 0, representing the absence of the feature at that location. This is achieved by first exporting T and F as RGB images in GIS. Then, these RGB images are *converted* to grayscale images to transform from a 3-layer RGB image to a single-layer grayscale image by averaging Red, Green, and Blue pixel values. Finally, to *convert* the grayscale images to binary raster files, pixel values from the range 0 to 255 are quantized to binary values 0 and 1, using appropriate thresholding methods [28]. After this process, we get 2-dimensional binary matrices T^r and F^r. Therefore, $t_i^r[u,v] = 1$ ($f_k^r[u,v] = 1$) represents the existence of the i–th transmission line (k–th feature), while $t_i^r[u,v] = 0$ ($f_k^r[u,v] = 0$) represents the absence of it at location (u,v) of the image. An example of extracting the feature f_1="roads" from F and transforming it into a binary raster file f_1^r is shown in Figure 3.3.

3.2.2 Parameterization

In this step, the methodology finds a set of parameters, P, to describe and compare route options based on C. It compute these parameters using T^r and F^r from the previous step. Each parameter $p_{ij} \in P$ describes a route option t_i^r for a decision consideration c_j. For example, a parameter p_{31} is used to represent the cable length for the route option t_3^r and the decision consideration c_1 = "length of cable needed". Thus:

$$P = \{p_{ij} : p_{ij} = \lambda_j(t_i^r, c_j, F^r), 1 \leq i \leq |T^r|, 1 \leq j \leq |C|\} \tag{3.4}$$

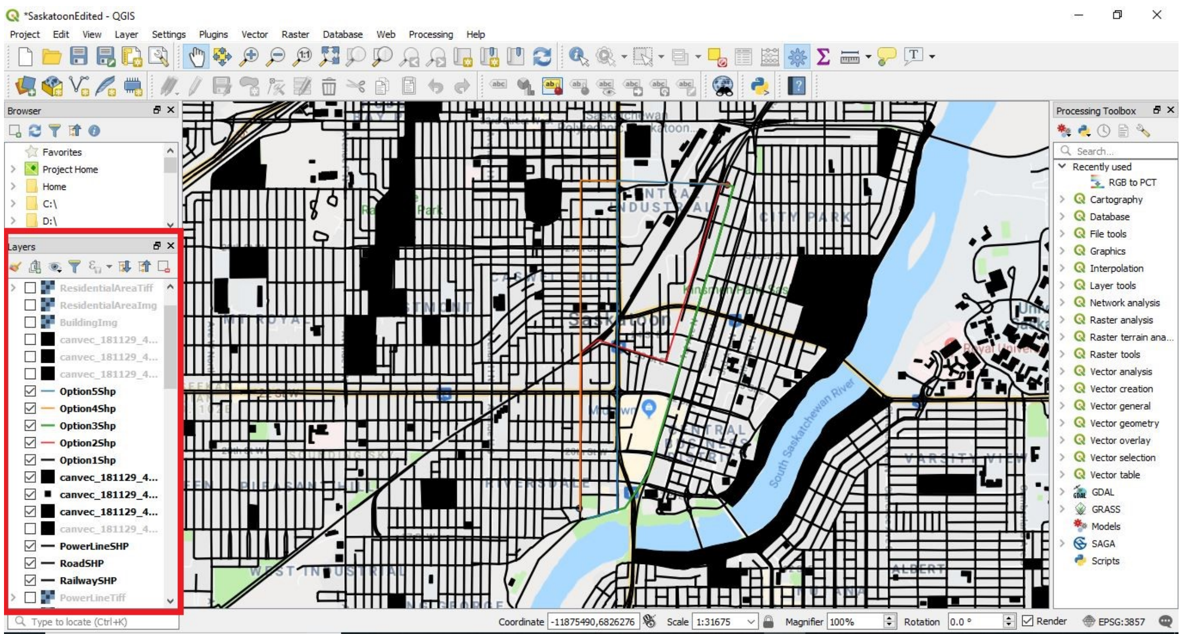


Figure 3.2: An example of extracting features of interest by *choosing* a subset of features from the "Layer" toolbox in QGIS.

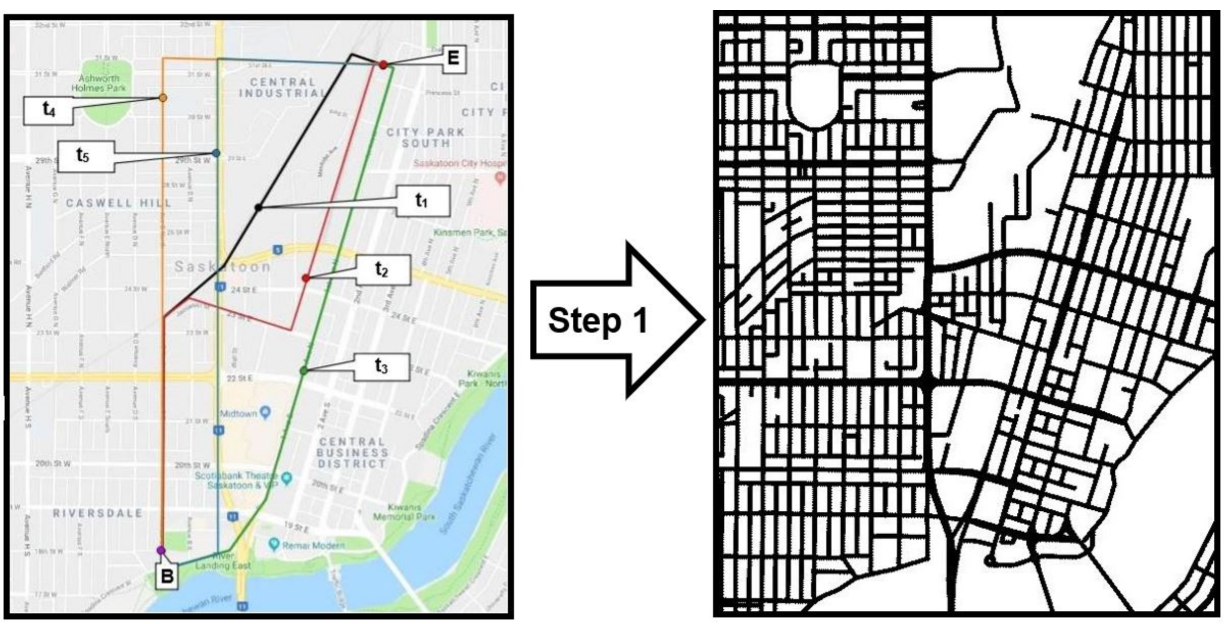


Figure 3.3: Extracting f_1 from F and transforming it to f_1^r.

where $\lambda_j()$ is the parameterization procedure for the decision consideration c_j.

Most of the parameters that reflect the decision considerations in set C can be *computed* using information such as length of the path, length of the path through feature k, length of the path along the feature k, and/or number of intersections with feature k. They may be used in isolation or in combination to get parameter values that best reflect the decision consideration. The way to evaluate parameter values depends on the type of the decision consideration. To *compute* a parameter value of t_i^r for a decision consideration c_j which depends on the length of its path, the methodology simply counts the number of pixels with value 1 in t_i^r. We then multiply the count with the pixel scale L – the actual length on earth's surface per pixel – to find the length of the route t_i^r. This can be expressed as follows:

$$p_{ij} = \left(\sum_{u=1,v=1}^{u=U,v=V} t_i^r[u,v] \right) * L. \tag{3.5}$$

On the other hand, to *compute* a parameter value of t_i^r for a decision consideration c_j which depends on the length of its path through a feature f_k^r, the methodology counts the number of pixels with value 1 common to both t_i^r and f_k^r and multiply the count by L. This process can be expressed as follows:

$$p_{ij} = \left(\sum_{u=1,v=1}^{u=U,v=V} \left(t_i^r[u,v] \mathbf{AND} f_k^r[u,v] \right) \right) * L \tag{3.6}$$

where **AND** is an extended version of the logical AND operation. **AND** is a bit-wise logical AND operation performed on corresponding pixel values of the two raster images. It is important to skeletonize the image using a thinning algorithm before evaluating the length to make sure that the path in the image is one-pixel-wide line; otherwise, the length evaluated will not be accurate.

Lastly, to *compute* a parameter value of t_i^r for a decision consideration c_j, which depends on the percentage of the path that coexists with a feature f_k^r, the methodology uses a combination of Equations (3.5) and (3.6) as follows:

$$p_{ij} = \frac{\sum_{u=1,v=1}^{u=U,v=V} \left\{ t_i^r[u,v] \mathbf{AND} f_k^r[u,v] \right\}}{\sum_{u=1,v=1}^{u=U,v=V} t_i^r[u,v]} * 100\%. \tag{3.7}$$

This step is repeated for all route options in T^r. To compare the route options $\{t_i^r\}_{i=1}^n$ for the specific decision consideration c_j, the methodology simply compares the values of p_{ij} for $1 \leq i \leq n$.

It is to be noted that the parameterization step is not limited to computing parameters that are directly related to length. It can be extended to accommodate more types of parameters that may not seem obvious at first. For example, the decision makers may want to compare the route options based on the time required to build a transmission line on those routes. A parameter that describes the duration of time needed to build a transmission line on a route option can be estimated using industrial rules of thumb and combinations of Equation 3.6. Here, the parameterization step uses the industrial rule of thumb for the average time required to build a transmission line per unit length on a residential area, commercial area, or any area represented by a feature f_k^r. For instance, if we want to estimate the time needed to build a line on a route option t_i^r that passes through residential and forest areas, we can find the length through these two features and multiply the lengths by the industrial rules of thumb for duration– the average time needed to build unit length of transmission line on these areas.

Similarly, the parameterization step can use industrial rules of thumb for cost – average cost of building a line per unit length over different features – to compute a parameter that describes the estimated cost of building a transmission line on a route option. Simple feature modifications can also be used to find a parameter that describes the length of path in close vicinity of a feature and not directly on it. For example, if we want the length of a path that is within x meters of an existing transmission line, the parameterization step can edit the raster file representing the utility lines and increase the thickness of its lines to x meters. It can then use the modified-thickened feature to compute the length of path in the vicinity of existing transmission line. Thus, the model is flexible enough to accommodate various types of parameters that resonates with more types of decision considerations.

3.2.3 Score evaluation

Herein, the methodology obtains a set of cumulative scores A for all route options using both P and W as inputs. This is done in two sub-steps. First, it scores every route option for each decision consideration to get the set of scores S. A score $s_{ij} \in S$ represents a score of t_i^r for a decision consideration c_j. Then, it uses the WSM method on S and W to compute the set A. The two sub-steps are expressed as follows:

$$S = \{s_{ij} : s_{ij} = \phi_j(p_{ij}), 1 \leq i \leq |T^r|, 1 \leq j \leq |W|\}, \tag{3.8}$$

$$A = \{a_i : a_i = \sum_{j=1}^{|W|} s_{ij} * w_j, 1 \leq i \leq |T^r|\} \tag{3.9}$$

where $\phi_j()$ is the scoring procedure for the decision considerations c_j.

To compare all routes $\{t_i^r\}_{i=1}^n$ for a decision consideration c_j, the methodology simply compares the elements in the subset of P such that $\{p_{ij}\}_{i=1}^n$. Successively, the scores $\{s_{ij}\}_{i=1}^n$ are evaluated for routes $\{t_i^r\}_{i=1}^n$ and the decision consideration c_j using the scoring procedure $\phi_j()$. This is repeated for all decision considerations in the set C. Ideally, the work wants to *score* on a scale of 1 to 5, where a score of 5 is the best and score of 1 is the worst. The industrial partner in this research mostly uses this range to score different route options, however, it can differ for different considerations depending on the decision makers' requirements. Thus, the scoring procedure is subject to recommendations of the decision makers, and needs manual tuning accordingly. For example, one way to score transmission lines $\{t_i^r\}_{i=1}^n$ for a consideration c_j on a scale of 1 to N is by finding the range between the maximum and minimum values in the sub-set $\{p_{ij}\}_{i=1}^n$ and then dividing the range into N equal sub-ranges. If the decision considerations c_j is an opportunity, then the highest score, N, is assigned to the highest sub-range and lower scores are assigned to the lower sub-range in descending order. Thus, the lowest sub-range is assigned a score of 1. On the contrary, if the decision considerations c_j is a constraint, then the highest score, N, is assigned to the lowest sub-range and lower scores are assigned to the higher sub-range. Thus, in this example, a transmission line route option t_i gets a score depending on which sub-range its parameter value p_{ij} falls under and whether the decision consideration c_j is a constraint or an opportunity.

Once all the route options in T^r are scored for every decision consideration in C, the methodology applies the WSM method to compute the set of cumulative scores A. This is achieved by multiplying the scores in the set S with weights in the set W. This step is simply a dot product of S with W.

3.2.4 Ranking route options

In this step, the methodology simply sorts T^r in descending order based on the values of the set A. Figure 3.4 and 3.5 summarizes the four steps of the methodology. In the next section, the work shows in detail how to apply the methodology.

Step 1: Extract and Transform
Input: T, F // route options and features in raster or vector format
Output: T^r, F^r // route options and extracted features in raster format
Procedure 1: $\xi(\ )$ // extracts necessary features

1. Choose y features of interest from F
2. Save y features in F^e

Procedure 2: $\eta(\ )$ // transforms inputs to binary raster files

1. **For** every f_k in F^e
2.   Convert f_k to RGB Image
3.   Convert RGB image to grayscale image
4.   Convert grayscale image to Binary Raster file f_k^r
5. end **For**
6. $F^r = \{f_k^r\}_{k=1}^{y}$
7. Repeat Procedure 2 for T to produce T^r

Step 2: Parameterize
Input: T^r, F^r, C //route options and features in raster format
Output: P // parameters describing characteristics of route options
Procedure 3: $\lambda_j(\ )$ // computes the parameters needed to describe route options

1. **For** every c_j in C
2.   **For** every t_i^r in T^r
3.   Compute parameter value p_{ij} using equations (3.5), (3.6), or (3.7)
4.   end **For**
5. end **For**

Figure 3.4: The pseudo-code for step 1 and step 2 of the proposed methodology.

Step 3: Score
Input: P, W // parameters describing route options and weights of importance for considerations
Output: A // cumulative scores of route options
Procedure 4: $\phi_j(\ )$ // scores a route based on parameters

1. **For** every p_{ij} in P
2. Score t_i for c_j as s_{ij}
3. end **For**

Procedure 5: $S \cdot W$

1. Compute A as dot product of S and W

Step 4: Rank
Input: A // cumulative scores of route options
Output: Ranked list of T // an array indexes of route options sorted in descending order based on A
Procedure 6: Sort

1. Sort A in descending order
2. Store T according to the order of A

Figure 3.5: The pseudo-code for step 3 and step 4 of the proposed methodology.

3.3 Implementation

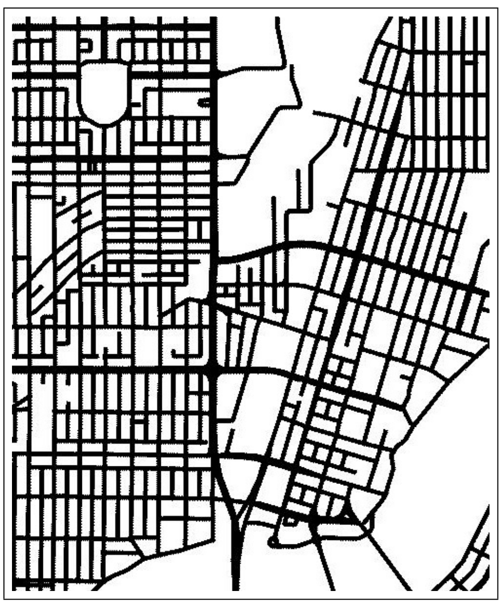
In section implements the proposed methodology to rank a given set of route options for a transmission line in Saskatoon, SK, Canada that is shown in Figure 1.2. For this real scenario, the four input sets are defined as follows:

$$\begin{aligned}
&T = \{t_1, t_2, t_3, t_4, t_5\},\\
&F = \{f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”},\\
&\quad f_5 = \text{“}residential\ area\text{”}, f_6 = \text{“}commercial\ area\text{”}\},\\
&C = \{c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}\}, and\\
&W = \{w_1 = 5,\ w_2 = 3\}.
\end{aligned}$$

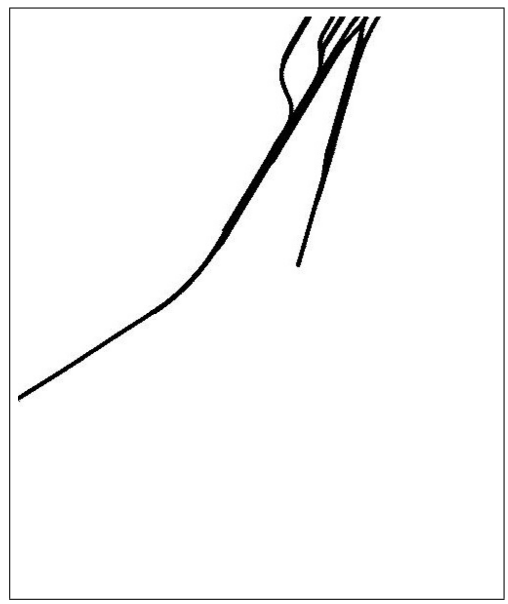
In the first step, the main objective is to obtain the set of extracted features F^r and the set of route options T^r in raster format from the input sets F and T. Given the two decision considerations in the set C, only features f_1 = “*roads*” and f_2 = “*railways*” have impacts on the route options. Therefore, $F^e = \{f_1, f_2\}$. The sets F^e and T are then converted to raster files as described in the methodology such that $F^r = \{f_2^r, f_2^r\}$ and $T^r = \{t_1^r, t_2^r, t_3^r, t_4^r, t_5^r\}$. These steps are performed in QGIS desktop software version 3.2.3. The geographic area of interest is 4×5 Km^2, and the image resolution is 1000×1250 pixels. Therefore, the pixel scale L is 4 m/pixel. The extracted and transformed features are shown in Figures 3.6(a) and 3.6(b), respectively, and the five route options to be ranked are shown in Figures 3.7(a) through 3.7(e) in raster format.

The second step *computes* the parameters needed to describe and differentiate the route options for the decision considerations in C. The parameterization step is performed using MATLAB R2018a. It takes T^r, F^r, and C as inputs and outputs the set P. As there are two decision considerations in C, two parameters p_{i1} and p_{i2} are evaluated for every route option t_i^r, $1 \leq i \leq 5$. Thus, for this scenario, $P = \{p_{11}, p_{21}, p_{31}, p_{41}, p_{51}, p_{12}, p_{22}, p_{32}, p_{42}, p_{52}\}$. The parameter values $\{p_{i1}\}_{i=1}^{5}$ are evaluated using Equation (3.5) because the decision consideration c_1 is the cable length needed, and this directly relates to the length of the route. For the parameters $\{p_{i2}\}_{i=1}^{5}$, we use a modified version of the Equation (3.7) because c_2 is the “ROW used” which depends on the percentage of path that coexists with the two features: roads and railways. The modified Equation (3.10) is given by:

$$p_{i2} = \frac{\sum_{u=1,v=1}^{u=U,v=V} \Big\{ t_i^r[u,v] \mathbf{AND} \Big(f_1^r[u,v] \mathbf{OR} f_2^r[u,v] \Big) \Big\}}{\sum_{u=1,v=1}^{u=U,v=V} t_i^r[u,v]} 100\%,\ 1 \leq i \leq 5 \tag{3.10}$$



(a)



(b)

Figure 3.6: The extracted features in F^r in binary raster format where parts (a) and (b) show roads and railways, respectively.

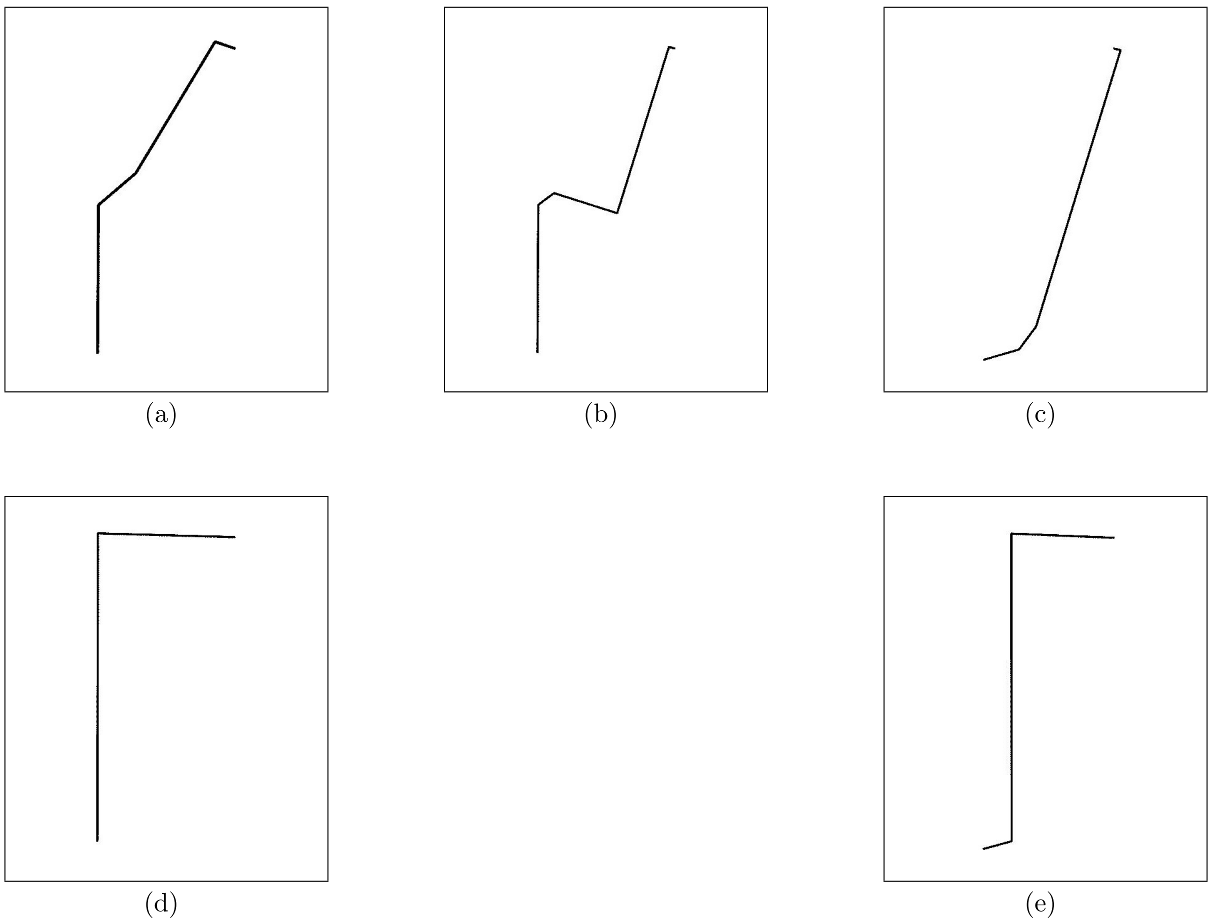


Figure 3.7: The five route options in T^r in binary raster format where parts (a), (b), (c), (d), and (e) show t_1^r, t_2^r, t_3^r, t_4^r, and t_5^r, respectively.

where **OR** is an extended version of the logical OR operation, U=1250 and V=1000. Here, **OR** is a bit-wise logical OR operation performed on corresponding pixel values of the two raster images. The parameters computed in Step 2 are shown in Table 3.2.

Table 3.2: The values for the parameters p_{i1} and p_{i2}.

Options	p_{i1} **(Km)**	p_{i2} **(%)**
t_1	2.1440	96.2687
t_2	2.5360	81.2303
t_3	1.9520	95.9016
t_4	2.8360	83.3568
t_5	2.7920	82.9513

The third step *computes* the set of cumulative scores A based on the parameter values in P and weights W. To score the five route options for the decision consideration c_1, the methodology simply compares the parameters $p_{11}, p_{21}, p_{31}, p_{41}$, and p_{51} and scores them on a scale from 1 to 5 as $s_{11}, s_{21}, s_{31}, s_{41}$, and s_{51}, respectively. Similarly, to score the same five route options for the decision consideration c_2, it compares the parameters $p_{12}, p_{22}, p_{32}, p_{42}$, and p_{52} and scores them as $s_{12}, s_{22}, s_{32}, s_{42}$, and s_{52}, respectively. The set of scores S generated by the automated method for this scenario is shown in Table 3.3 under columns "s_{i1}" and "s_{i2}" with the label "auto" (automated). Then, it finds the set of cumulative scores A using the WSM method as shown in Table 3.3 under the column "a_i" with the label "auto".

In the last step, the methodology ranks the route options based on the set of cumulative scores A. This step takes A as an input and outputs the ranked list of T.

3.4 Results and Discussion

To evaluate the effectiveness of the methodology, the work compares the outputs of the automated method ("auto") with the manual method ("man"). The results obtained using the manual method are provided by our industry partner, and they are shown in Table 3.3 with the label "man". Due to a Non-Disclosure Agreement (NDA) between the industry partner and our research group, I am unable to disclose the company name and the data it shared with us. Henceforth, the company will be referred as "industry partner" and its data will be referred as "industrial data" or "manual data" interchangeably. The results from the two methods are also illustrated in Figure 3.8.

Table 3.3: The scores generated by the proposed methodology and manual method.

Options	s_{i1}		s_{i2}		a_i	
	auto	**man**	**auto**	**man**	**auto**	**man**
t_1	2	2	5	5	25	25
t_2	1	2	1	3	8	19
t_3	2	2	5	5	25	25
t_4	1	2	1	1	8	13
t_5	1	1	1	2	8	11

Figure 3.8 shows that the methodology does agree with the manual method in ranking t_1 and t_3 as the first and second most optimal paths, respectively. However, the proposed methodology does not give any preference to the route options t_2, t_4, or t_5, and it ranks them equally likely, while the manual method does rank t_2, t_4 and t_5 as the third, fourth, and fifth optimal options, respectively. This discrepancy could have resulted from the two ways of scoring the route options. In the manual method, the scores are roughly evaluated using human experience and intuition, which is subjective. In contrast, the employed scoring procedure in our methodology uses crisp logic by dividing the range of each set of parameters into equal intervals. Nevertheless, given the obtained results, it is evident that our methodology shows a similar pattern in ranking the route options, compared to the manual method. The average error in computing the length and ROW used are 7% and 4.2%, respectively. The errors are mainly due to the skeletonization process used for thinning the routes, scaling the path on the map, and the scoring procedure. For example, in the process of reducing the thickness, the thinning procedure also reduces the length of the path slightly which is unwanted. In addition, when the proposed methodology multiplies the number of pixels by L to find the length of the path scaled to what is on earth, it multiplies the number of pixels with the distance represented by the width of the pixel. This introduces some errors when the path is diagonal because pixels are square in shape and the diagonal distance is longer than their sides. Thus, the accuracy of the methodology may increase by increasing the resolution of the images.

Overall, testing the methodology with a small set of inputs proved the feasibility and practicality of the methodology. To validate the methodology thoroughly, it needs to be tested with larger input sets. Due to limited availability of real projects' data, synthetic

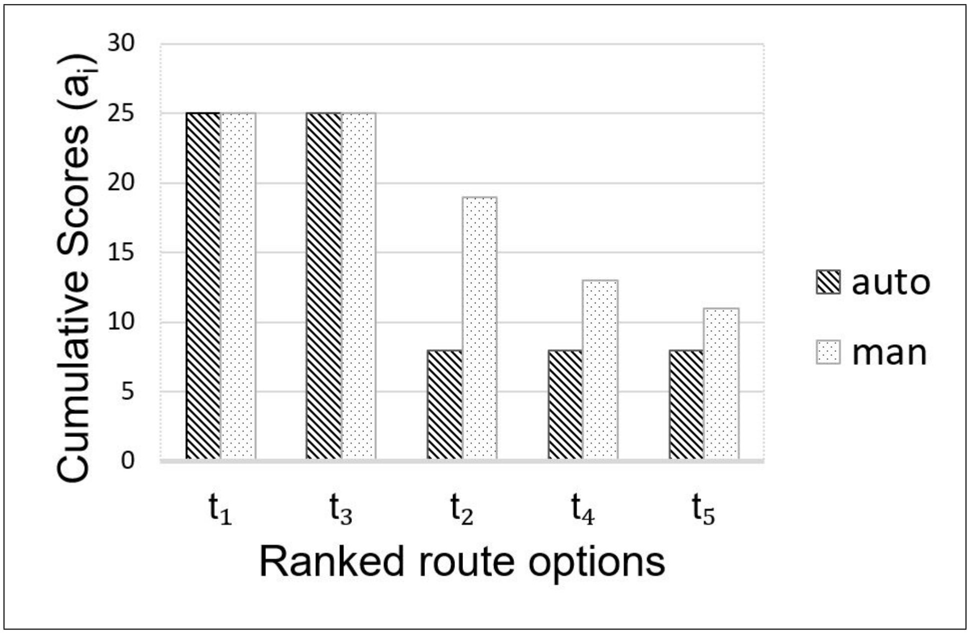


Figure 3.8: Cumulative scores generated by the proposed methodology and the manual method.

data are generated to represent hypothetical scenarios. Next chapter uses two methods to generate synthetic data. Later in Chapter 5, the synthetic data will be used to validate the proposed methodology further.

Chapter 4

Generating synthetic data

In the previous chapter, the proposed methodology was validated with only five route options and two decision considerations. To further validate the automated ranking process, it is important to test the methodology with more route options and decision considerations. Due to limited industrial and geographic data, it is necessary to synthesize hypothetical routes and features to test the methodology with more test cases. The process of generating hypothetical routes and features are described in the following sections. In the next chapter, the synthetic data will be used in different combinations to thoroughly validate the proposed methodology.

4.1 Generating random routes

A set of random routes can be generated from a given route using a perturbation process. Section 4.1.1 introduces the perturbation process that will be used in this work and section 4.1.2 describes it in detail. Later, section 4.1.3 demonstrates the process by generating 150 new random routes. Table 4.1 lists the symbols used in the perturbation process. These 150 random routes will be used in Chapter 5 to validate the automated ranking process in depth.

4.1.1 Perturbation

Perturbation is a deviation from its regular or normal path or state caused by an outside influence. This work uses perturbation process to generate many paths from a given path by

Table 4.1: Nomenclature for the perturbation process

Symbol	Description
T^r	set of all manual route options in raster format
n	number of route options in T, $n = \|T\|$
t_i^r	a manual route option in the set T, $1 \leq i \leq n$
U	number of rows in t_i^r
V	number of columns in t_i^r
B	coordinates of starting point of T_i^r representing source
E	coordinates of ending point of T_i^r representing sink
TT^r	a set of randomly generated route options in raster format
o	number of randomly generated route options in TT^r, $o = \|TT^r\|$
tt_g^r	a randomly generated route option in the set TT^r, $1 \leq g \leq o$
R_g	an intermediate point through which tt_g^r must pass through
l_g	a straight line with user defined slope α and y-intercept β
α	represents the slope of l_g
β	represents the y-intercept of l_g
ST^r	a set of randomly generated route options in raster format after smoothing operation
st_g^r	a randomly generated route option in the set ST^r, $1 \leq g \leq o$
κ	number of samples from tt_g^r to generate st_g^r

influencing it to traverse through intermediate points before reaching a sink from a source. Here, the source is a generating sub-station and the sink is a load center represented by B and E, respectively.

Let us consider a given route $t_i^r \in T^r$ that starts from B and ends at E. We want to generate a set of new random routes $TT^r = \{tt_g^r\}_{g=1}^{o}$, where $o = |TT^r|$, using the given t_i^r. In order for the new routes to have some variance, they are forced to pass through intermediate points. Let R_g be an intermediate point selected randomly for a new route tt_g^r. To have some control over where intermediate point must lie on the area of interest, let us draw a straight line l_g and restrict the intermediate point R_g on it. The straight line l_g can be controlled by changing its slope α and y-intercept β. This is illustrated in Figure 4.1 where l_1 and l_2 have the same slope but different y-intercept values. On the other hand, l_1 and l_3 have the different slopes.

Figure 4.1 shows an example where a set of random routes $TT^r = \{tt_1^r, tt_2^r, tt_3^r\}$ is generated from a single manual route option t_1^r. The intermediate points through which the routes tt_1^r, tt_2^r, and tt_3^r had to traverse through are R_1, R_2, and R_3, respectively. It is evident from the figure that by randomly choosing intermediates points, the proposed perturbation process can bring a lot of variation in the new generated routes. This is important because we do not want similar routes, rather, various possible routes. If we are just testing similar routes, we will not be exploring other possible paths that could be better options. We can also see that by changing the y-intercept and the slope of the line, we can have some control over where the random point is selected. The perturbation process to generate random routes from a given route can be summarized as follows:

1. Identify the end points of the given route t_i^r on the map M and mark them as source B and sink E.

2. Draw a straight line l_g between the source and sink for the given slope α and y-intercept β.

3. Randomly choose a point R_g on l_g.

4. *Trace* a path tt_g^r from B to E such that it passes through the intermediate point R_g.

5. *Smooth* tt_g^r using sampling method to generate final output st_g^r

6. Repeat steps 3 to 5 for as many new routes needed. Change the slope and the y-intercept of l_g to get some controlled variation in the generated routes.

The procedures to *Trace* and *Smooth* a route will be discussed in the next subsection.

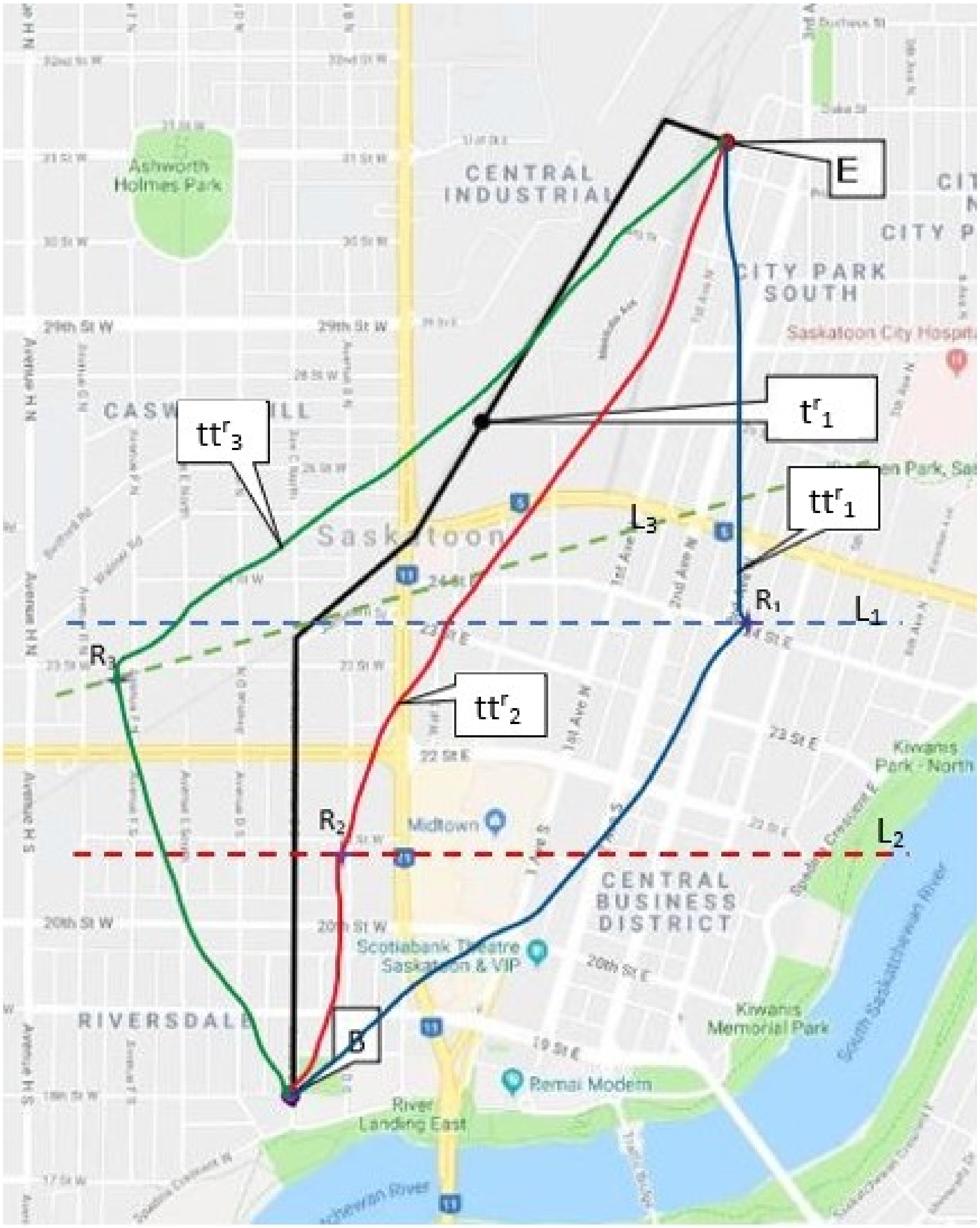


Figure 4.1: Generating random routes using the perturbation process.

4.1.2 Procedures

The proposed perturbation process consists of three procedures: *routePerturbation*, *routeTracer*, and *routeSmoother*. The first procedure, *routePerturbation*, is the main procedure that calls the other two procedures as its sub-routines. Thus, the inputs to the *routePerturbation* procedure are the inputs to the whole perturbation process.

routePerturbation

The fundamental function of this procedure is to take a sample path and generate a set of random paths from it. In order to achieve this, it takes five inputs, namely t_i^r, o, β, α, and κ. Here, t_i^r is a sample manual route in binary raster format. This can be represented as a 2-dimensional binary matrix with U rows and V columns. Thus, $t_i^r(u,v) = 1$ represents the presence of the route and $t_i^r(u,v) = 0$ represents the absence of it at (u,v) location, where $1 \leq u \leq U$ and $1 \leq v \leq V$. o is an integer that tells the *routePerturbation* how many random routes to generate. The α and β are integer inputs that represents the slope and y-intecept of l_g. These parameters give control to the users to manipulate the straight line l_g. Lastly, the κ is actually an input for the *routeSmoother* procedure that is called in this routine. The *routePerturbation* procedure takes these inputs and generates the set of random routes ST^r by executing the following steps.

Firstly, *routePerturbation* finds the coordinates of the end points for t_i^r. These can be done by checking the condition $\sum(neighboring\ 8\ pixels'\ values) < 1$ for a pixel representing the presence of the route. If true, it is an end point; else, it is not an endpoint. The coordinates of the endpoints are stored as B and E. Secondly, it uses the user inputs, α and β, to find the equation of a line l_g between B and E. It then randomly chooses a point on the line and stores the coordinates as R_g. Thirdly, it calls the *routeTracer* to trace a path from B to E via R_g. Here, tracing a path means setting $tt_g^r(u,v) = 1$ for coordinates where the path exists. At first, it calls *routeTracer* to trace a path from B to R_g, and then from R_g to E. The two traces are stored in the same raster file tt_g^r using the extended **OR** operation. The extended **OR** operation is a modified logical OR operation performed on two raster files. It is like a bit-wise OR operation performed on the corresponding pixels of two raster files that are at the same location. This helps to overwrite a raster file with a new raster file without loosing any data from the previous file. Lastly, it calls the *routeSmoother* procedure to smooth out rough edges of the path produced by *routeTracer*. These steps generate a route from B to E by detouring through R_g. The third and forth steps are repeated o times to produce required set of random routes ST^r. The pseudo-code for *routePerturbation* is shown in Figure 4.2.

routeTracer

This procedure is a sub-routine of *routePerturbation* and is used to draw a route between two points on the map M. It takes two coordinates S and D representing the source and destination of the path, respectively. It produces an output Z^r representing a traced route between S and D in the raster format. It starts from S by initializing a temporary point $P = S$. Here, P represents the current position on the map. The procedure traces a path by setting the binary pixels of Z^r to 1 for every position P travels. As for a 2-D raster map, every position has 8 neighboring positions around it. Thus, from any position, P has 8 possible directions to travel at any instance. The next probable position for P is represented by PP. In order to avoid a straight line from S to D, PP is selected randomly. However, to guide the path towards destination D, PP is restricted to positions that either reduces or equals the current distance to D from P. If the randomly chosen PP satisfies these conditions, P is set to PP. This assignment traces a path from P to PP. These steps are repeated until P reaches the neighboring positions of D. The pseudo-code for routeTracer is provided in Figure 4.3.

routeSmoother

This procedure is also a sub-routine of *routePerturbation* and is used to smooth out the route traced by *routeTracer*. It takes tt^r_g and κ as inputs and outputs a smoother version of tt^r_g represented by st^r_g. This is important because because *routeTracer* produces a path that has a lot of irregular rough turns due to the degree of freedom given to *routeTracer*. However, sharp turns are generally undesirable in routes for electrical transmission line. Since, we want to represent probable routes for electrical transmission lines, the rough route produced by *routeTracer* is smoothened in this procedure. It takes κ samples along the rough route tt^r_g and stores it as an ordered list of coordinates. It then traces straight lines between the points in the ordered list. Thus, by increasing or decreasing the number of samples κ, we can control the smoothness of the route st^r_g. The pseudo-code for the procedure is summarized in Figure 4.4. The next subsection shows in detail how to apply the procedures.

4.1.3 Implementation

This subsection uses the proposed perturbation procedures to generate random routes from a given route. For test scenario 1, the given route is t^r_1 in raster format as shown in Figure

Procedure: routePerturbation()
Inputs:
$t_i^r : a\ route\ in\ raster\ format,$
$o : number\ of\ routes\ to\ generate,$
$\alpha : slope,$
$\beta : y - intercept,$
$\kappa : number\ of\ samples.$
Output: ST^r // a set of randomly generated route options in raster format.

1. Let B be the coordinates of start point of t_i^r
2. Let E be the coordinates of end point of t_i^r
3. Let l_g be the stright line represented by the slope α and y-intercept β
4. $[U, V]$ = size(t_i^r) //returns the number of rows and columns of the t_i^r raster file
5. $ST^r = \{\}$ //instantiates a blank set of random routes
6. Instantiate tt_g^r as a blank raster image with U rows and V columns and set all pixel values to 0.
7. **For** g = 1:o
8.   Randomly choose a point R_g on the line l_g
9. $tt_g^r = tt_g^r$ **OR** routeTracer(B, R_g)
10. $tt_g^r = tt_g^r$ **OR** routeTracer(R_g, E)
11. st_g^r = routeSmoother(tt_g^r, κ)
12. $ST^r = ST^r \cup \{\text{st}_g^r\}$
13. End **For**
14. Return ST^r

Figure 4.2: The pseudo-code for routePerturbation.

Procedure: routeTracer()
Inputs: S, D. //S is the start point and D is the end point
Output: Z^r. //Z^r is a route in the raster format

1. $P = S$ //Let P be a temporary point and instantiate it as start point S.
2. Let PP be a probable next point
3. Instantiate Z^r as a blank raster image with U rows and V columns and set all pixel values to 0.
4. **While** ($P \notin neighbor\ of\ D$)
5. d_1 = distance from P to D
6. Randomly choose a neighboring position PP from 8 possible positions around P.
7. d_2 = distance from PP to D
8. **If** ($d_2 \leq d_1$)
9. **Then**
10. Set the position PP on Z^r to 1
11. $P = PP$
12. **Else** go back to step 4
13. End **While**
14. Return Z^r

Figure 4.3: The pseudo-code for routeTracer.

Procedure: routeSmoother()
Inputs: tt^r_g, κ // model route option and integer representing number of samples to take
Output: st^r_g //set of smoothened random routes

1. sRoute = Skeletonize(tt^r_g) //make tt^r_g one-pixel-wide line

2. Take κ samples equidistant from each other along the length of sRoute and store their coordinates in an ordered list

3. st^r_g = trace straight lines between the points in the ordered list of coordinates by setting pixels along the line to 1

4. Return st^r_g

Figure 4.4: The pseudo-code for routeSmoother.

3.7(a). Let $o = 5$, $\alpha = 0^o$, $\beta = U/2$, and $\kappa = 10$. The random routes generated by the perturbation process for the above inputs are shown in Figure 4.5.

Figure 4.5(a) through Figure 4.5(e) shows the individual routes generated for the input provided. As $o = 5$, the perturbation process generates 5 routes. Figure 4.5(f) shows all the routes in one image for better visualization. From Figure 4.5(f) we can see that all the routes generated are different. The difference in paths are mainly due to the intermediate points they had to pass through. It is also evident in Figure 4.5(f) that the intermediate points for the routes generated lay on a horizontal line approximately at the center of the image. This is because the slope $\alpha = 0^o$ and y-intercept $\beta = U/2$ making l_g a straight horizontal line passing through the center of the image.

The work uses the perturbation procedure described above to generate more random routes for different input scenarios. Table 4.2 shows all the input sets used to generate different random routes for this research. To showcase the variations in routes produced using various input sets, Figure 4.6 illustrates the routes generated by the first six scenarios from Table 4.2. Here, every figure, Figure 4.6(a) to Figure 4.6(f), shows the five routes generated by the perturbation process. In total, 150 routes are generated using 30 scenarios listed in Table 4.2. They are shown in Figure 4.7. These 150 routes along with the 5 manual routes will be used to validate our ranking methodology in the next section.

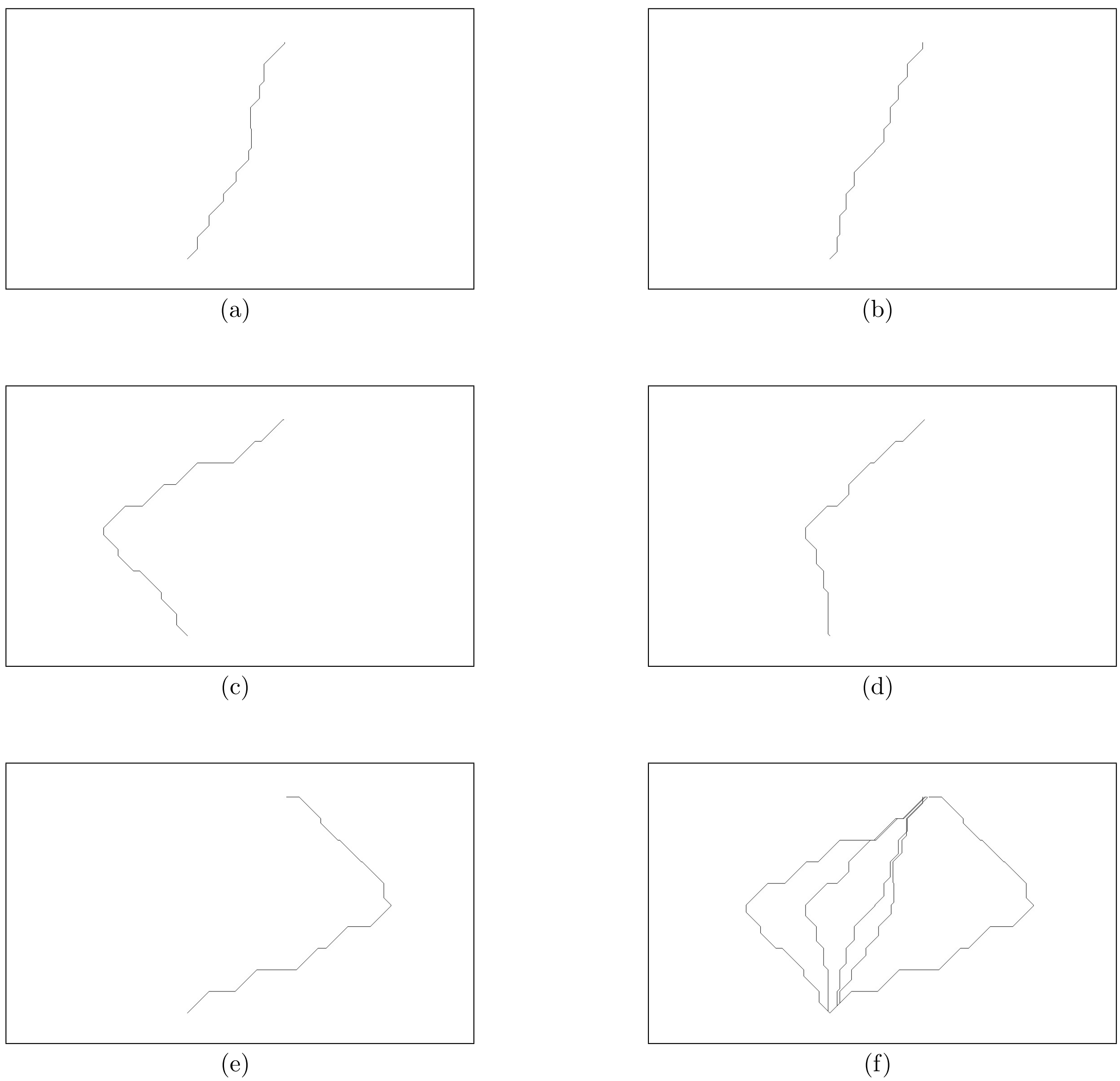


Figure 4.5: The five randomly generated routes in TT^r in binary raster format where parts (a), (b), (c), (d), and (e) show tt_1^r, tt_2^r, tt_3^r, tt_4^r, and tt_5^r, respectively. All the five generated routes in TT^r is shown in part (f).

Table 4.2: Implementing perturbation process for 30 different input scenarios.

Scenario	**input route**	o	α	β	κ
1	t_1^r	5	0^o	$U/2$	10
2		5	30^o	$U/2$	10
3		5	60^o	$U/2$	10
4		5	120^o	$U/2$	10
5		5	150^o	$U/2$	10
6		5	180^o	$U/2$	10
7	t_2^r	5	0^o	$U/10$	8
8		5	0^o	$U/5$	8
9		5	0^o	$2 * U/5$	8
10		5	0^o	$U/2$	8
11		5	0^o	$3 * U/5$	8
12		5	0^o	$7 * U/10$	8
13	t_3^r	5	0^o	$U/2$	6
14		5	0^o	$U/2$	8
15		5	0^o	$U/2$	10
16		5	0^o	$U/2$	12
17		5	0^o	$U/2$	14
18		5	0^o	$U/2$	16
19	t_4^r	5	0^o	$U/10$	10
20		5	30^o	$U/5$	10
21		5	60^o	$2 * U/5$	10
22		5	120^o	$U/2$	10
23		5	150^o	$3 * U/5$	10
24		5	180^o	$7 * U/10$	10
25	t_5^r	5	0^o	$7 * U/10$	10
26		5	30^o	$3 * U/5$	12
27		5	60^o	$U/2$	12
28		5	120^o	$2 * U/5$	12
29		5	150^o	$U/5$	12
30		5	180^o	$U/10$	12
	Total	150			

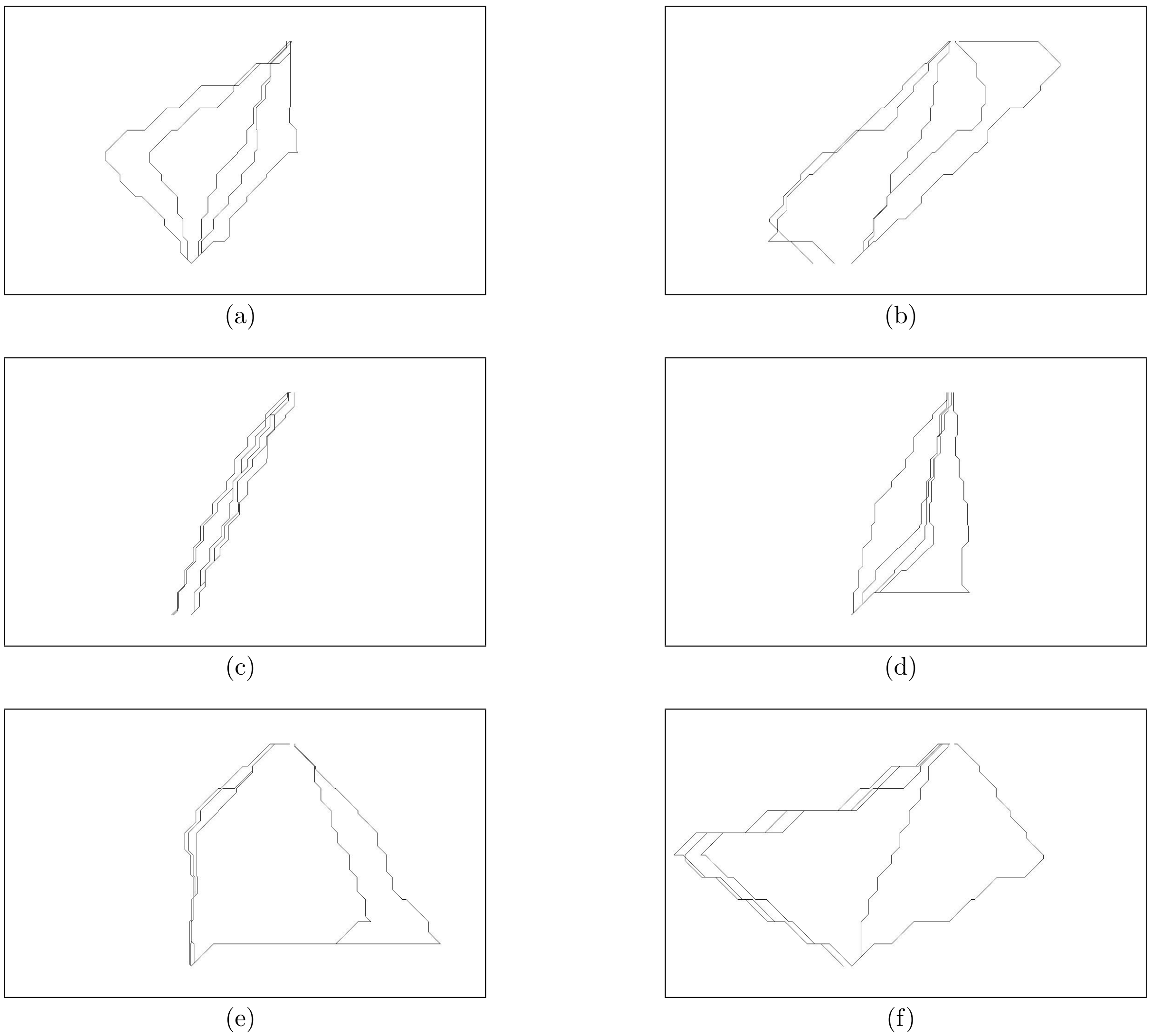


Figure 4.6: The first thirty randomly generated routes using the perturbation process where parts (a), (b), (c), (d), (e), and (f) show the five routes generated by scenarios 1, 2, 3, 4, 5, and 6, respectively.

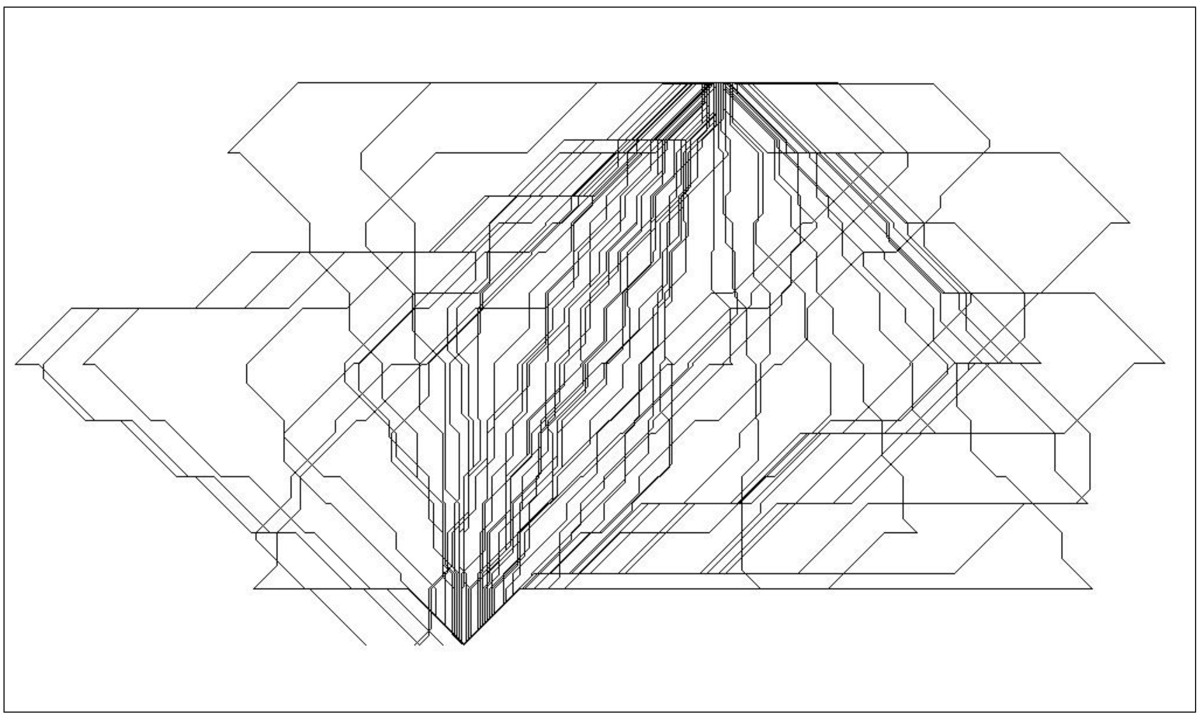


Figure 4.7: The 150 routes generated by the perturbation process using the 30 scenarios listed in Table 4.2.

4.2 Generating features for more considerations

The case study in this thesis is on a geographic area located at Saskatoon, SK, Canada. The area has limited features, namely roads, railways, rivers, parks, private land, and commercial area. To test the methodology for more decision considerations, hypothetical features need to be generated. Thus, this section proposes a *featureGenertor* procedure that can generate hypothetical features. As the man-made and natural features are represented by binary raster files in this thesis, the *featureGenertor* procedure uses simple image processing techniques to generate new hypothetical features in the same format.

4.2.1 Procedure

Table 4.3: Nomenclature for feature generation process

Symbol	**Description**
F^r	set of extracted natural and man-made features in raster format
y	number of extracted features in F^r, $y = \|F^r\|$
f_k^r	a feature in the set F^r, $1 \le k \le y$
U	number of rows in f_k^r
V	number of columns in f_k^r
γ	coordinates of a vertex of the rectangular feature
τ	coordinates of a vertex diagonal to γ
FF^r	a set of generated features in raster format
θ	number of generated features in FF^r, $\theta = \|FF^r\|$
ff_ρ^r	a generated feature in the set FF^r, $1 \le \rho \le \theta$

Features can be represented by geometric shapes such as lines, triangles, rectangles, circles, ellipses, and polygons. To simplify the problem, the *featureGenertor* generates rectangles on the map to represent hypothetical features by taking the coordinates of diagonal vertices of the rectangles. This simplification can be justified by the fact that rectangles can be used to represent many types of features on a geographic area. For example, if the sides of a rectangle are the same, it represents a square feature, and if horizontal sides of the rectangle are very small, it can represent a vertical line and vice versa. Thus, a *featureGenertor* that can generate rectangular features can also be used to generate features represented by points, lines, squares, and rectangles. Moreover, by generating multiple rectangular shapes, it can be used to represent complex shapes too. However, this procedure can also be extended to accommodate other forms of geometric shapes easily. For example, it can be modified to generate a circle by taking the coordinates of the center point and the radius, or to generate a triangle by taking the coordinates of the three vertices. The next sub-section describes a procedure, *featureGenerator*, to generate rectangular features. Table 4.3 shows the symbols used in the *featureGenerator* procedure.

featureGenerator

The fundamental function of this procedure is to take a sample feature and coordinates of two points to generate a rectangular feature in raster format. In order to achieve this, it takes three inputs: f_k^r, γ, and τ. Here, f_k^r is a sample feature of the map M in binary raster format. γ and τ are the coordinates of the diagonal vertices of a rectangle. The user defines the location and size of the rectangle with the help of these two points. The procedure takes these inputs and generates the rectangle such that it lies between x-coordinates and y-coordinates of the two points, γ and τ. It outputs the hypothetical feature as ff_ρ^r, which is also a binary raster file. This can be achieved by assigning the pixels under the rectangle to 1, while, assigning all other pixels in the layer to 0. For example, Figure 4.8 shows a hypothetical feature "restricted area" generated using the *featureGenerator* procedure for the diagonal vertices γ and τ. Restricted area can be any geographic area that needs to be avoided by an electrical transmission line. The pseudo-code for the proposed procedure is shown in Figure 4.9.

4.2.2 Implementation

To test the proper functionality of the proposed procedure, two hypothetical rectangular areas are generated, one on the left and the other on the right region of the map M. The proposed *featureGenerator* produces these two hypothetical features shown in Figure 4.10. They will be used in the next chapter to extensively validate the ranking method proposed in Chapter 3.

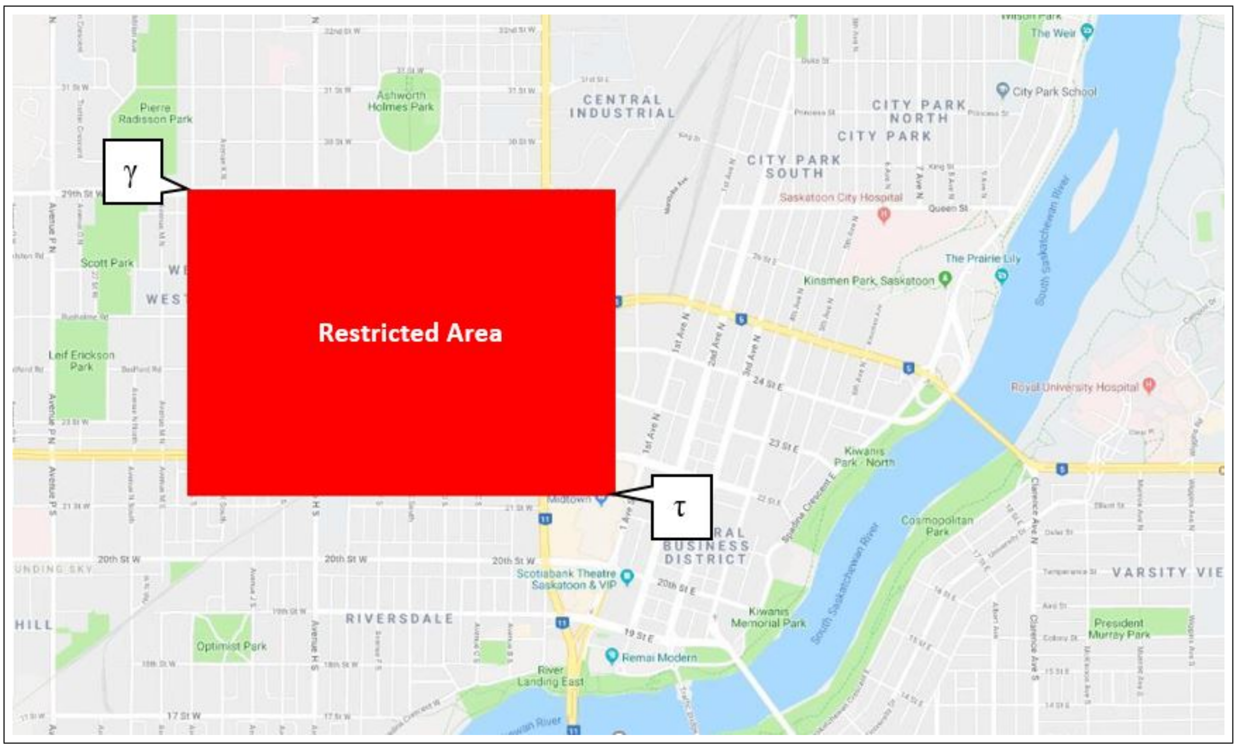


Figure 4.8: Rectangular feature representing "restricted area" with diagonal vertices γ and τ.

Procedure: featureGenerator()
Inputs: f_k^r, γ, τ
Output: ff_ρ^r

1. Let P be a point on a feature where (P_x, P_y) are the (x, y) coordinates of P.
2. Let $\gamma_x \leq$ X $\leq \tau_x$.
3. Let $\gamma_y \leq$ Y $\leq \tau_y$.
4. $[U, V]$ = size(f_k^r) // returns the number of rows and columns of the f_k^r raster file
5. **For** P_y = 1: V
6.   **For** P_x =1 : U
7.   **If** (P_x in X & P_y in Y)
8. ff_ρ^r (P_x,P_y) = 1
9.   **Else**
10. ff_ρ^r (P_x,P_y) = 0
11.   End **For**
12. End **For**
13. Return ff_ρ^r

Figure 4.9: The pseudo-code for featureGenerator.

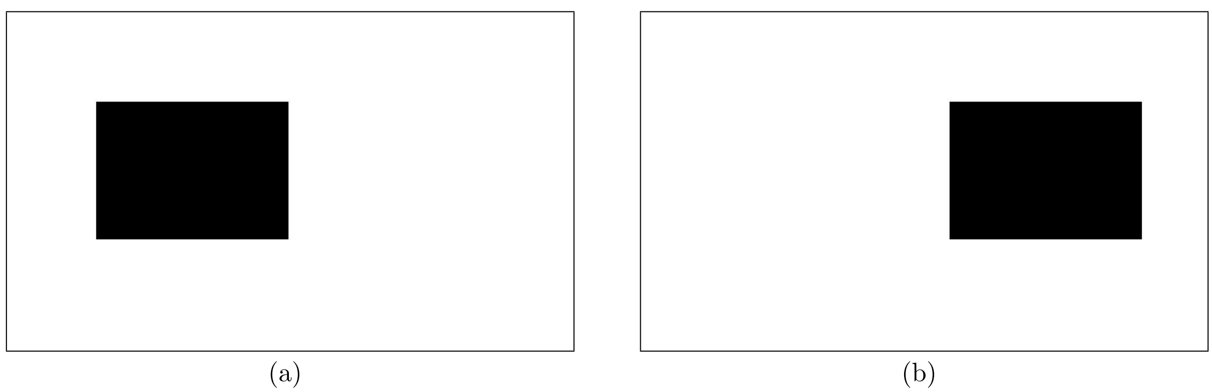


Figure 4.10: The features generated in FF^r in binary raster format where parts (a) and (b) show *restricted area 1* on the left and *restricted area 2* on the right regions, respectively.

Chapter 5

Validation

Previously, in Chapter 3, the thesis validated the proposed methodology for ranking of routes for electrical transmission lines with only five manual routes and two features by comparing it to industrial results. In this section, the work further verifies the methodology by testing it with synthesized data and comparing its output with predicted results. The methods used to generate synthetic data are discussed in detail in Chapter 4. The following sections will test the methodology for different input sets.

5.1 Testing with more route options in T

To check the methodology with more route options, the perturbation process is used, as discussed in section 4.1.3, to generate 150 random routes using 30 scenarios listed in Table 4.2. These random routes, shown in Figure 4.7, along with the 5 manual routes, shown in Figure 3.7 are given as the input set T to the proposed ranking process. The other inputs to the ranking process are the same as discussed in section 4.1.3. As the 5 manual routes are selected by humans after a thorough feasibility study, they are expected to rank top among all the route options in T. This is because the random routes do not try to avoid areas with constraints or utilize opportunities. The inputs and outputs for this test scenario are discussed in the following sub-sections.

5.1.1 Input

Proposed ranking methodology using WSM is tested with the following input sets:

$$\begin{aligned}
&T = \{t_1, t_2, t_3, ..., t_{155}\},\\
&F = \{f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”},\\
&\quad\quad f_5 = \text{“}residential\ area\text{”}, f_6 = \text{“}commercial\ area\text{”}\},\\
&C = \{c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}\}, and\\
&W = \{w_1 = 5,\ w_2 = 3\}.
\end{aligned}$$

Here, t_1 to t_5 are the manual routes, and t_6 to t_{155} are the randomly generated routes using the perturbation process.

5.1.2 Output

The scores generated by the automated ranking process are illustrated by a bar chart in Figure 5.1 for better visualization. The horizontal axis shows the route options in T given as input and the vertical axis represents the cumulative scores in A the route options received by the proposed methodology. Since there are 155 route options in set T, there are equal number of vertical bars on the horizontal axis. The top 15 routes from Figure 5.1 are shown separately in Figure 5.2. It is clear from this separate figure that the manual routes, t_1 to t_5, are ranked as top routes with a score of 25. All the randomly generated routes are ranked lower with a score of 19 or less.

5.1.3 Discussion

We can see from Figures 5.1 and 5.2 that the five manual route options are ranked higher than any of the randomly generated routes by the automated method. To a great extent, the output produced verifies the proposed methodology as it resembles the expected output. However, it is to be noted that all the manual route options have the same cumulative score of 25 in Figure 5.2. This result is a deviation from the output in section 3.3 as it ranked t_3 and t_1 as the top options and t_2, t_4, and t_5 as the bottom options. The deviation was caused by the increase in the number of route options in the set T. To understand this phenomenon, we need to understand how the ranking method scores the routes. As the automated method scores each route option on a scale of 1 to 5 for a particular consideration, it is limited to these values to differentiate the route options

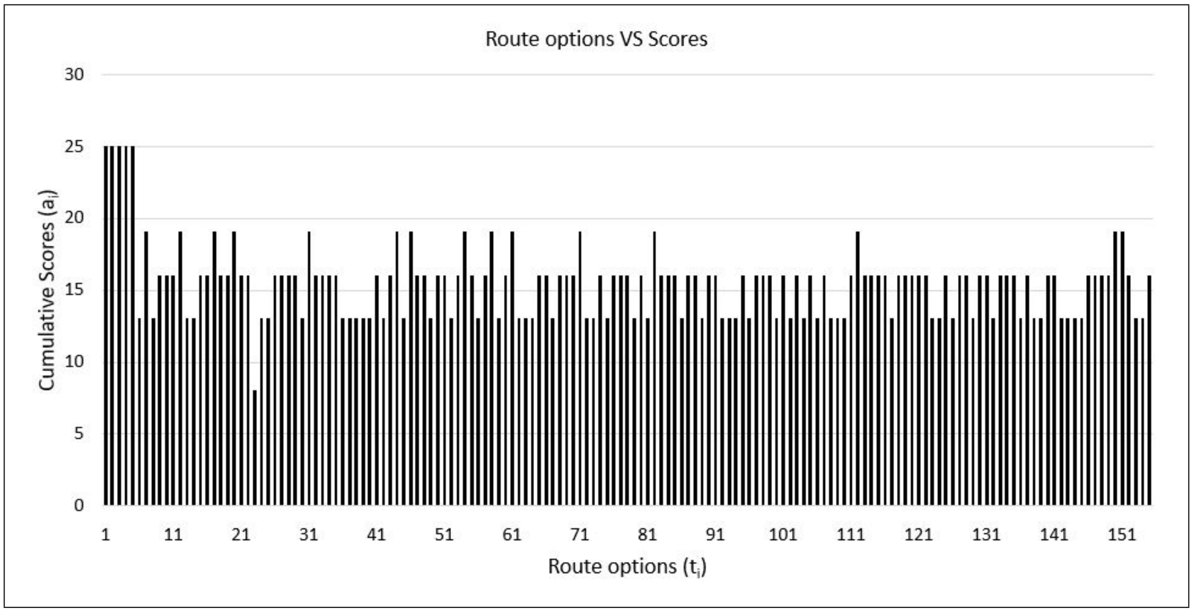


Figure 5.1: 155 route options and their cumulative scores according to the methodology using WSM.

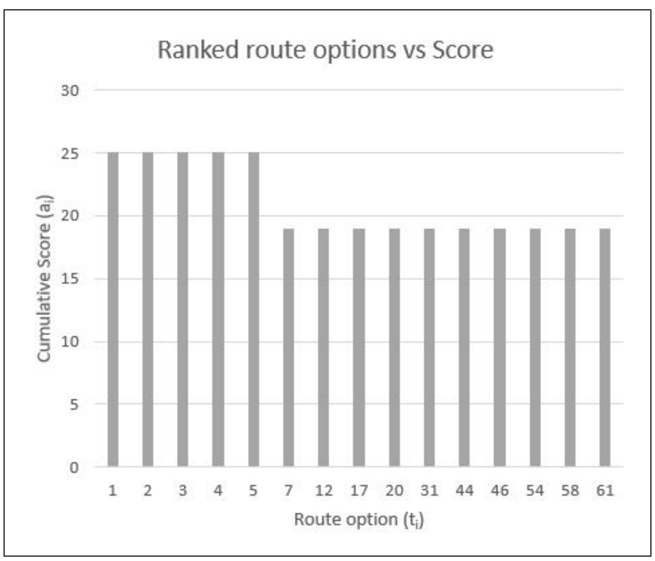


Figure 5.2: Top 15 route options and their cumulative scores according to the methodology using WSM.

amongst each other. It worked well with five route options. However, by increasing the number of routes, the variation of the parameter values increased by a great amount too. For example, with only 5 manual routes, the parameter values representing route's length varied from 1.95 km to 2.84 km. However, with 155 routes, the same parameter values varied from 1.95km to 4.2 km. As a result, a scale of 1 to 5 fell short to differentiate them. Thus, most of the manual routes got the same score despite having differences in their parameter values. This increase in variation could be because the randomly generated routes had parameter values with greater differences than the parameter values of manual routes themselves. A possible solution to this could be by scoring the route options on a scale larger than 1 to 5. For example, scoring on a scale of 1 to 100 instead of 1 to 5. This hypothesis is tested in the following section.

5.2 Testing with refined scoring scale

To confirm the hypothesis in the previous sub-section, the scoring procedure's scale in the proposed methodology is modified. Let us call the WSM with refined scale as Weighted Sum Model refined (WSMr) and let it score each route option on a scale of 1 to 100 instead of 1 to 5 for all the considerations. This should increase the resolution of scores and thus give more room for differentiating even minute differences in parameter values. The inputs and outputs for this scenario are discussed in the following sub-sections.

5.2.1 Input

The methodology using WSMr, which scores on a scale of 1 to 100 instead of 1 to 5, was tested for the same input sets as in section 5.1.1:

$$
\begin{aligned}
&T = \{t_1, t_2, t_3, ..., t_{155}\},\\
&F = \{f_1 = \text{``}roads\text{''}, f_2 = \text{``}railways\text{''}, f_3 = \text{``}rivers\text{''}, f_4 = \text{``}parks\text{''},\\
&\qquad f_5 = \text{``}residential\ area\text{''}, f_6 = \text{``}commercial\ area\text{''}\},\\
&C = \{c_1 = \text{``}cable\ length\text{''}, c_2 = ROW\ \ used\text{''}\}, and\\
&W = \{w_1 = 5,\ w_2 = 3\}.
\end{aligned}
$$

Here, t_1 to t_5 are the manual routes, and t_6 to t_{155} are the randomly generated routes using the perturbation process.

5.2.2 Output

The bar chart in Figure 5.3 displays the scores generated by the methodology using WSMr. The horizontal axis shows the route options in T given as input and the vertical axis represents the cumulative scores in A the route options received by the proposed methodology. There are 155 vertical bars on the horizontal axis, each representing the route options in T. Figure 5.4 focuses on the top 15 routes from Figure 5.3. The automated ranking process using WSMr also ranks the manual routes, t_1 to t_5, as top options.

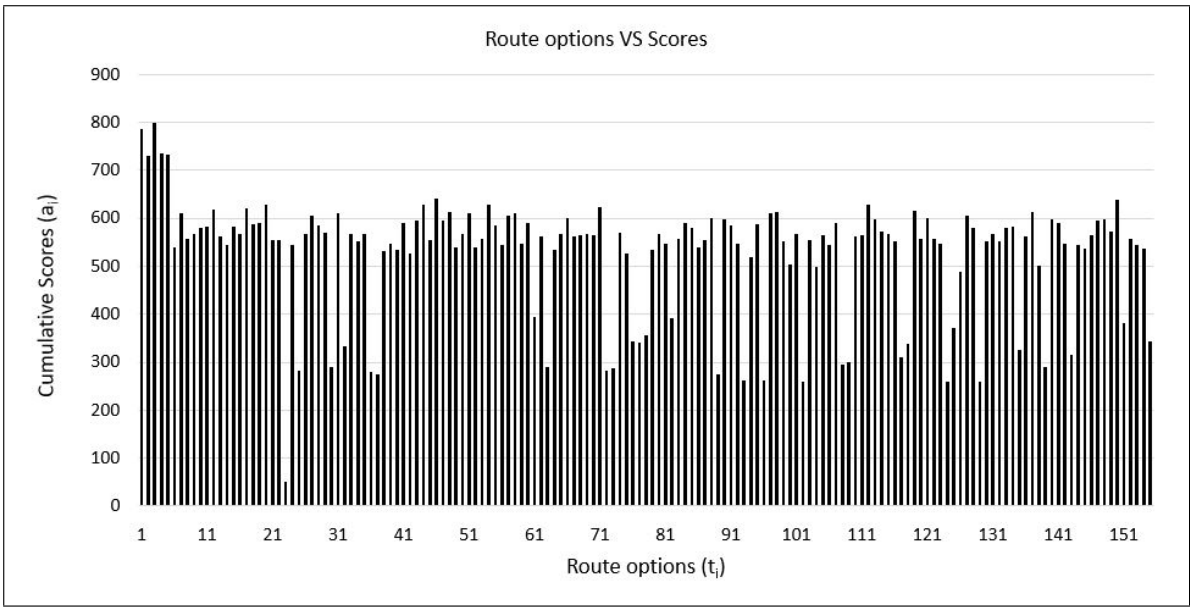


Figure 5.3: 155 route options and their cumulative scores according to the methodology using WSMr.

5.2.3 Discussion

We can see from Figure 5.4 that the five manual route options are ranked higher than any of the randomly generated routes by the methodology using WSMr too. Additionally, in this case, it successfully ranked t_3 and t_1 as top options among the manual routes. It is to be noted that the WSMr method has better capability in differentiating among route options. This is because the WSMr method scores every route option for each decision consideration from 1 to 100 in comparison to 1 to 5 in WSM. Thus, we can see that the

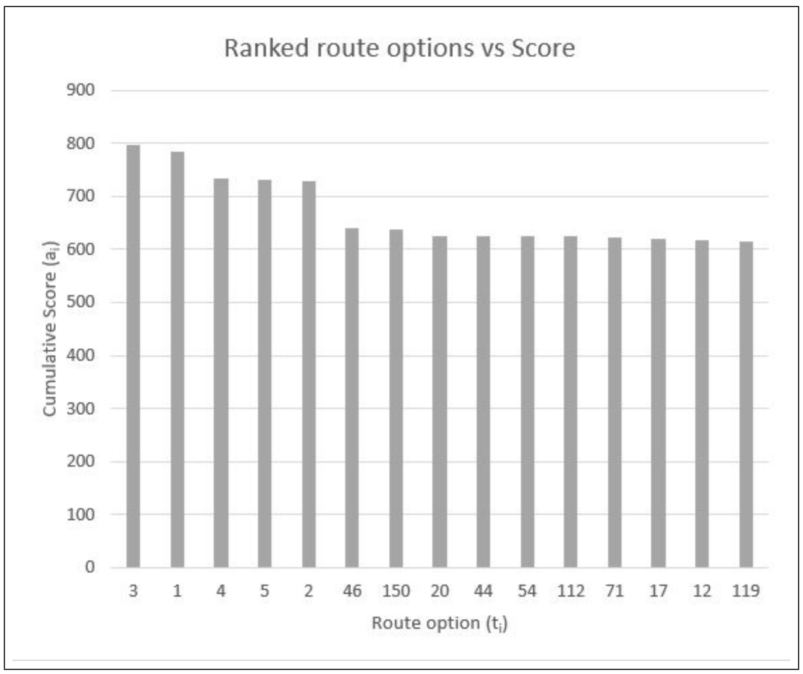


Figure 5.4: Top 15 route options and their cumulative scores according to the methodology using WSMr.

cumulative scores are from 1 to 800 as compared to 1 to 30 in the previous test in section 3.3. The maximum cumulative score possible is equal to the sum of product of the weights in W by the scale size N. As a result, the refined method has more numerical values to represent each route. This allows for much better ranking as opposed to the previous method.

5.3 Testing with a restricted area on the left

To check the accuracy of the methodology for additional considerations, this section tests the WSMr based methodology with an additional feature that represents restricted area on the left region of the map. This synthesized feature is generated using the *featureGenerator* procedure described in Chapter 4, section 4.2. The feature used is shown in Figure 4.10(a). As the restricted feature is on the left region of the map, it is expected that the top routes should lie on the right region of the map. The inputs and outputs for this scenario are discussed in the following sub-sections.

5.3.1 Input

The refined ranking methodology using WSMr, which scores on a scale of 1 to 100, is tested with an additional feature, "restricted area". Thus, the input sets are as follows:

$$T = \{t_1, t_2, t_3, ..., t_{155}\},$$
$$F = \{f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”},$$
$$f_5 = \text{“}residential\ area\text{”}, f_6 = \text{“}commercial\ area\text{”}, f_7 = \text{“}restricted\ area\text{”}\},$$
$$C = \{c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}, c_3 = \text{“}restricted\ area\ 1\ used\text{”}\}, and$$
$$W = \{w_1 = 5,\ w_2 = 3,\ w_3 = 10\}.$$

Here, t_1 to t_5 are the manual routes, and t_6 to t_{155} are the randomly generated routes using the perturbation process. Feature f_7 is a hypothetical restricted area and is generated using the *featureGenerator* procedure. The weight of importance for restricted area feature is $w_3 = 10$.

5.3.2 Output

The features extracted from F in the first step were $F^e = \{f_1, f_2, f_7\}$. f_7 was extracted and transformed because one of the decision considerations, c_3, is to minimize the use of restricted area. The output generated by the methodology using WSMr for the above input sets is shown in Figure 5.5. To check the proper functionality of the methodology, only the top 10 routes are shown in figure 5.6.

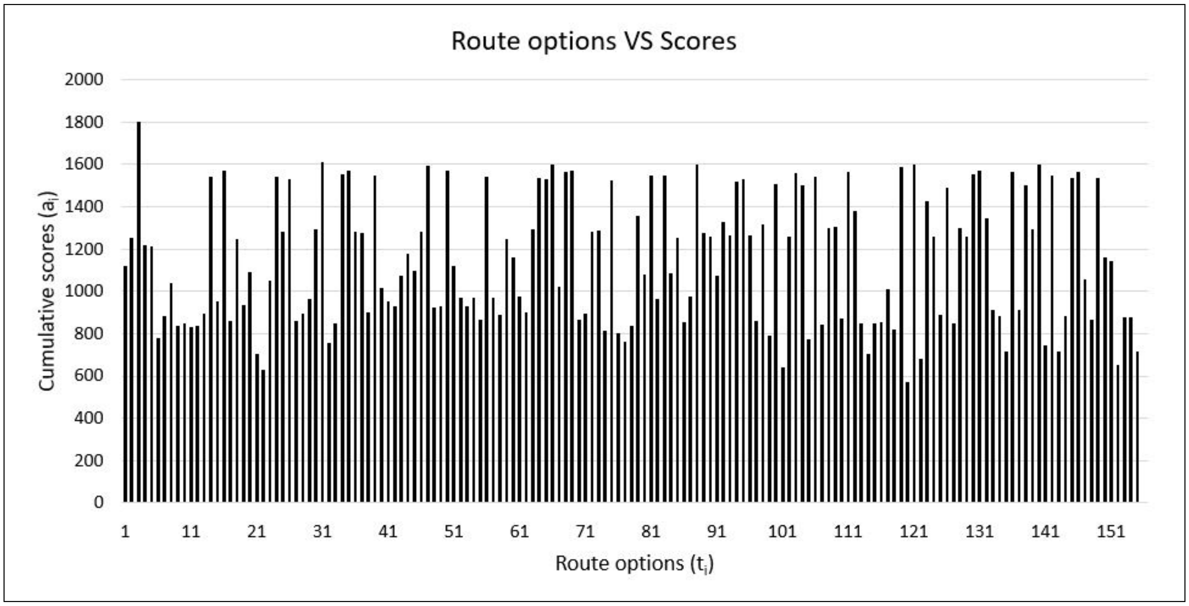


Figure 5.5: 155 route options and their cumulative scores according to the methodology using WSMr for an additional feature - restricted area on the left.

5.3.3 Discussion

We can see from Figure 5.5 that only one of the five manual route options, t_3, ranked top in this test scenario. In fact, all the other manual routes, are ranked lower than many randomly generated routes. This is because majority of the manual routes pass through the restricted area 1. As the weight of importance given to the consideration c_3 = "*restricted area* 1 *used*" is $w_3 = 10$, the other manual routes' cumulative scores are low. The cumulative scores are now from 1 to 1800 as compared to 1 to 800 in the previous test. This is because of the additional consideration and its higher weight. We can also see that the pattern of score from t_1 to t_{155} has also changed. The top 10 routes are illustrated in Figure 5.6. The manual routes are drawn with bolder lines. Figure 5.6 proves the proper functionality of the methodology because the automated method was successful in identifying the top routes that avoid the restricted area on the left region.

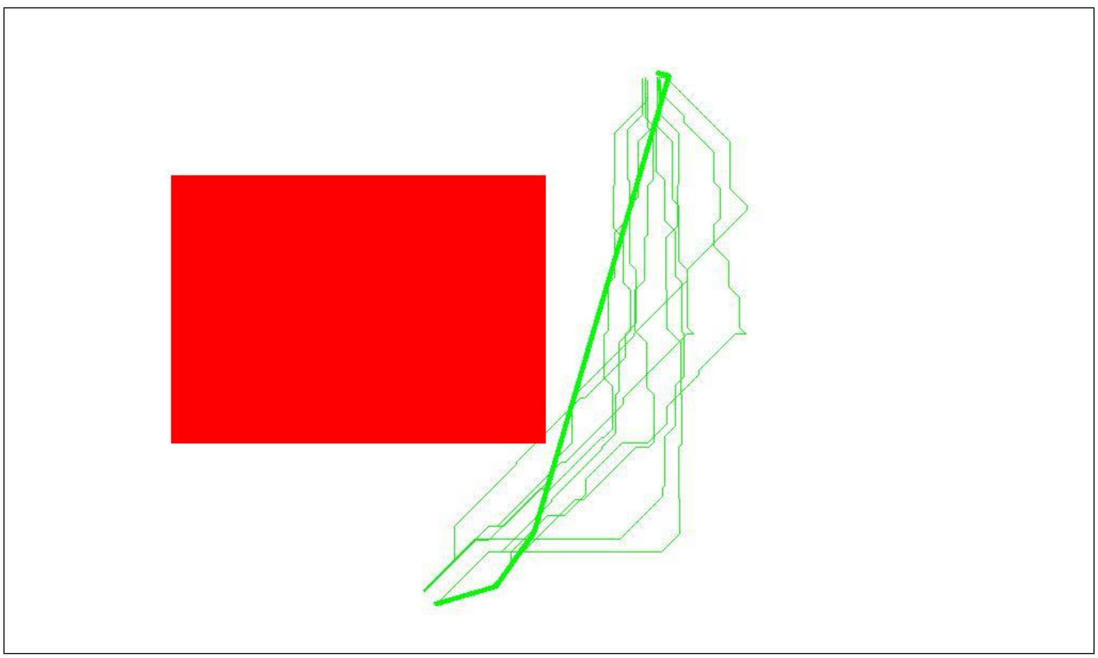


Figure 5.6: Top 10 route options according to the ranking method using WSMr for restricted area on the left region.

5.4 Testing with a restricted area on the right

Similarly, to confirm the accuracy of the methodology for features, this section tests it with a feature that represents restricted area on the right region of the map. This synthesized feature is also generated using the *featureGenerator* procedure described in chapter 4, section 4.2. The feature used is shown in Figure 4.10(b). As the restricted feature is on the right region of the map, it is expected that the top routes should lie on the left region of the map. The inputs and outputs for this scenario are discussed in the following sub-sections.

5.4.1 Input

The automated ranking method using WSMr, which scores on a scale of 1 to 100, was tested for a feature on the right region. The input sets are as follows:

$$
\begin{aligned}
&T = \{t_1, t_2, t_3, ..., t_{155}\},\\
&F = \{f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”},\\
&\qquad f_5 = \text{“}residential\ area\text{”}, f_6 = \text{“}commercial\ area\text{”}, f_7 = \text{“}restricted\ area\ 1\ used\text{”},\\
&\qquad f_8 = \text{“}restricted\ area\ 2\ used\text{”}\},\\
&C = \{c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}, c_3 = \text{“}restricted\ area\ 2\ used\text{”}\}, and\\
&W = \{w_1 = 5,\ w_2 = 3,\ w_3 = 10\}.
\end{aligned}
$$

Here, t_1 to t_5 are the manual routes, and t_6 to t_{155} are the randomly generated routes using the perturbation process. Features f_7 and f_8 are the hypothetical restricted area and is generated by the *featureGenerator* procedure.

5.4.2 Output

The features extracted from F in the first step were $F^e = \{f_1, f_2, f_8\}$. f_8 was extracted and transformed because one of the decision considerations, c_3, is to minimize the use of restricted area 2. The output generated by the methodology using WSMr for the above input sets is shown in Figure 5.7. To check the proper functionality of the methodology, only the top 10 routes are shown in figure 5.8.

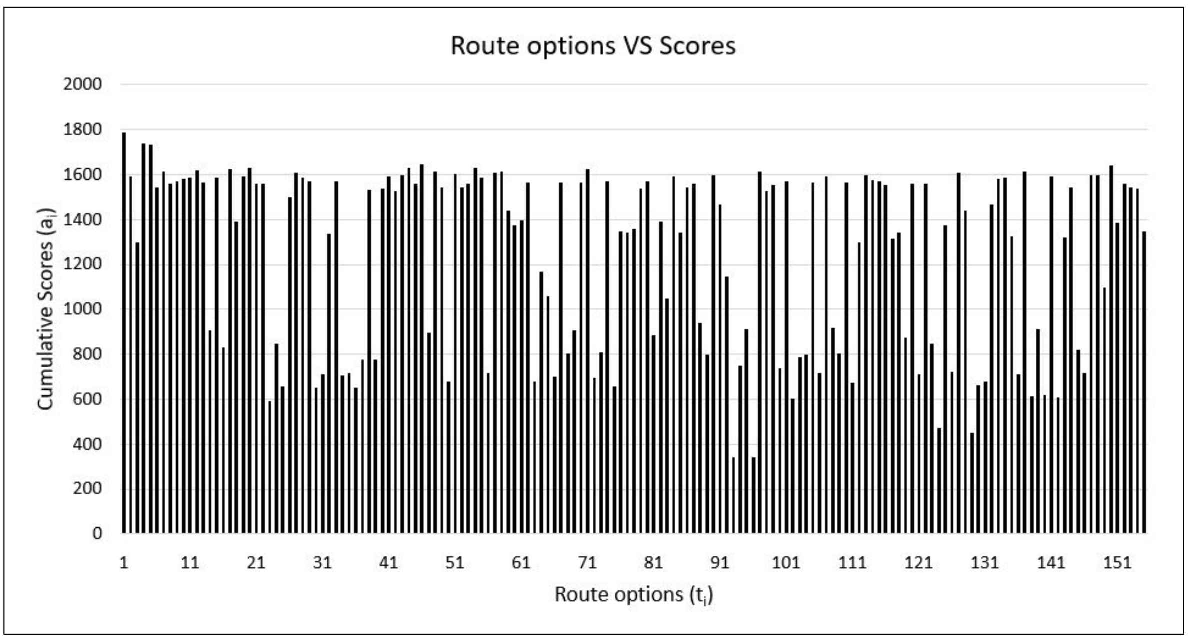


Figure 5.7: 155 route options and their cumulative scores according to the methodology using WSMr for an additional feature - restricted area on the right.

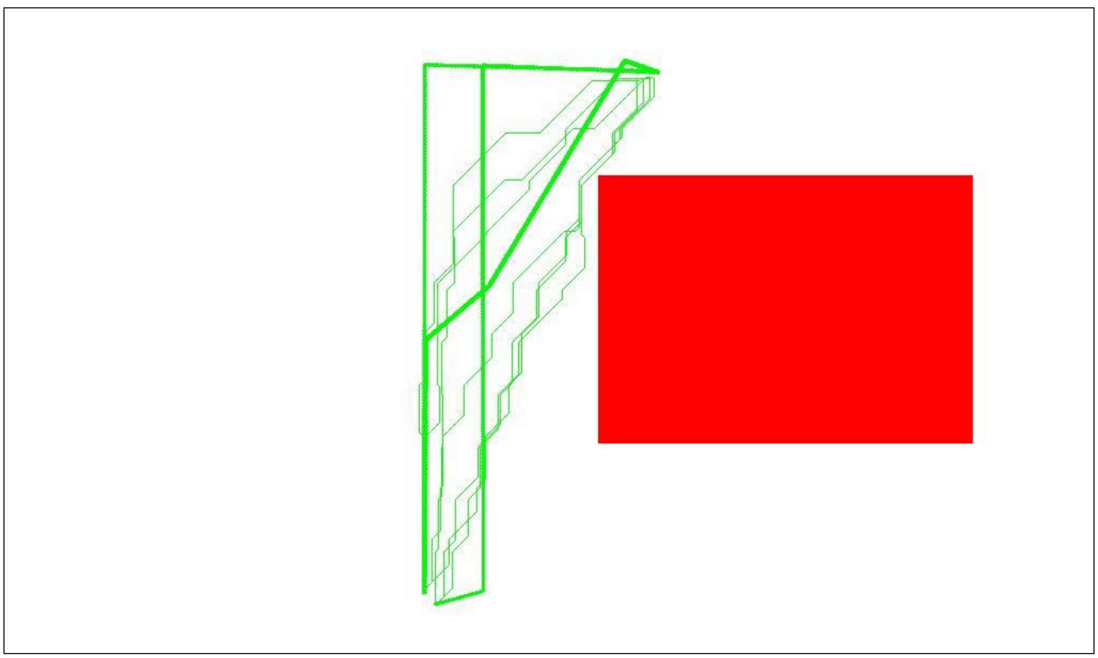


Figure 5.8: Top 10 route options according to the refined ranking method for restricted area on the right region.

5.4.3 Discussion

For this scenario, we can see from Figure 5.7 that three of the five manual route options, t_1, t_4, and t_5, are among the top routes. This is because majority of the manual routes does not pass through the restricted area 2. The cumulative scores are from 1 to 1800 because of the additional consideration and its higher weight. We can also see that the pattern of score from t_1 to t_{155} has also changed. The top 10 routes are illustrated in Figure 5.8. The manual routes are drawn with bolder lines. Figure 5.8 proves the proper functionality of the methodology because the automated method was successful in identifying the top routes that avoid the restricted area on the right region.

5.5 Testing with all manual and synthetic data

Finally, to prove the scalability and proper functionality of our methodology, it is tested with 155 routes and 6 decision considerations, which includes both manual and synthetic data. The synthetic data, routes and features, are generated using the *routePerturbation* and *featureGenerator* procedures described in Chapter 4. Depending on the decision considerations, the geographic features may be an opportunity or a constraint. The top routes ranked by the proposed methodology are expected to avoid the constraints and take advantage of the opportunities. The inputs and outputs for this scenario are discussed in the following sub-sections.

5.5.1 Input

Our proposed methodology using WSMr is tested with 155 route options and 6 decision considerations. The input set is as follows:

$$T = \{t_1, t_2, t_3, ..., t_{155}\},$$
$$F = \{f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”}, f_5 = \text{“}residential\ area\text{”},$$
$$f_6 = \text{“}commercial\ area\text{”}, f_7 = \text{“}restricted\ area\ 1\text{”}, f_8 = \text{“}restricted\ area\ 2\text{”},$$
$$f_9 = \text{“}private\ area\text{”}, f_{10} = \text{“}forest\ area\text{”}\},$$
$$C = \{c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}, c_3 = \text{“}restricted\ area\ 1\ used\text{”},$$
$$c_4 = \text{“}commercial\ area\ used\text{”}, c_5 = \text{“}private\ area\ used\text{”},$$
$$c_6 = \text{“}water\ area\ used\text{”}\}, and$$
$$W = \{w_1 = 5,\ w_2 = 3,\ w_3 = 10, w_4 = 8,\ w_5 = 8,\ w_6 = 8\}.$$

Here, t_1 to t_5 are the manual routes, and t_6 to t_{155} are the randomly generated routes using the perturbation process. Features f_7 and f_8 are hypothetical restricted areas and are generated by the *featureGenerator* procedure.

5.5.2 Output

For considerations c_1, we need the route options in T, where as for consideration c_2, we need features f_1 and f_2. Similarly, for considerations c_3, c_4, c_5, and c_6 we require features f_7, f_6, f_4, f_9, and f_3. Thus, in the first step, features extracted from F are $F^e = \{f_1, f_2, f_3, f_4, f_6, f_7, f_9\}$. The extracted features and transmission line route options are then transformed into raster files that are used in the parameterization step. The methodology uses WSMr in the third step to score the route options and sorts them in descending order according to their scores. The output generated by the methodology using WSMr for the above input sets is shown in Figure 5.9. To check the proper functionality of the methodology, only the top 10 routes are illustrated in Figure 5.10. In Figure 5.10, the red and yellow areas represent regions on the map with constraints and opportunities, respectively. The green lines are the top 10 routes options ranked by the automated ranking methodology using WSMr.

5.5.3 Discussion

We can see from Figure 5.10 that the top ranked routes successfully avoid the regions with higher constraints (areas in red) and utilize the corridors with opportunities (areas in yellow). None of the top 10 routes pass through undesirable areas, such as restricted area, water bodies, commercial areas, and private land, except one. This could be due to limited number of route options that fulfill all the decision criteria. Moreover, we can only see one manual route (in bold green lines) that was among the top 10 routes. Other manual route options ranked lower than many randomly generated routes as they crossed through restricted areas. This test proves that the proposed methodology can successfully identify routes that minimize areas with constraints and maximize areas with opportunity.

Till now, the research has successfully proposed, implemented, tested, and validated a methodology that can be used to automate the ranking of routes for electrical transmission lines. However, it still needs human expertise to quantitatively dictate the priorities of decision considerations using weights of importance W. Although this technique is widely used in the industry to indicate the importance of various decision considerations with respect to others, it may be inconvenient to many to come up with integer values to show

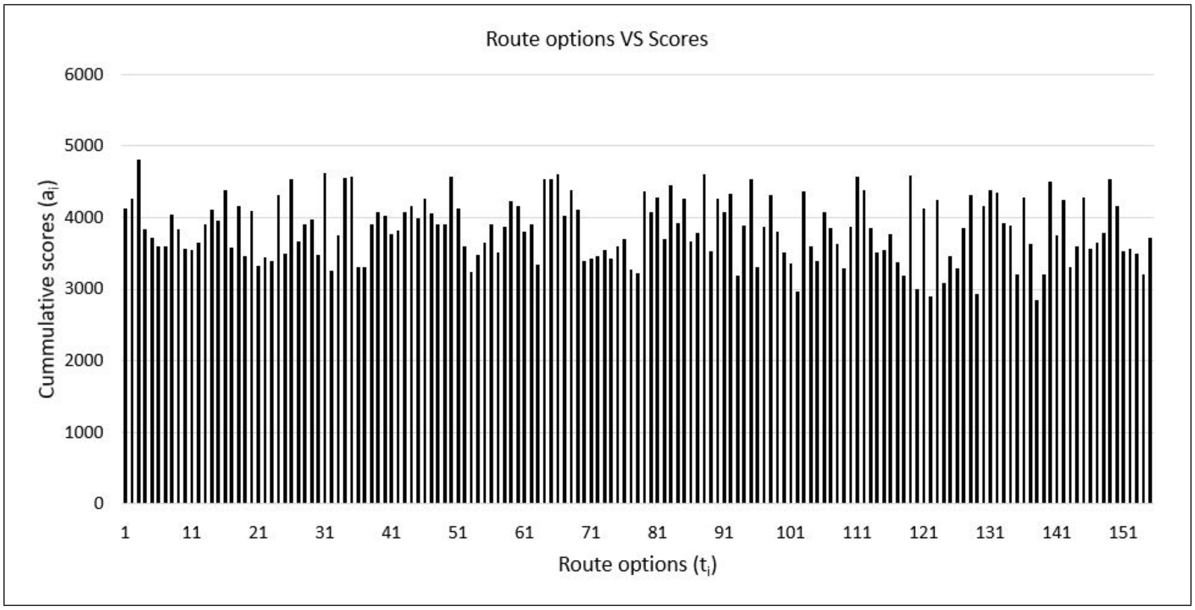


Figure 5.9: 155 route options and their cumulative scores according to the methodology using WSMr when tested for 6 decision considerations.

their preferences. Another approach to taking preferences of decision considerations can be using linguistic variables. Using day-to-day language to define the importance of considerations may make the methodology more user friendly and adoptable. To explore this idea, the thesis implements the same methodology using fuzzy logic in the next chapter. From hereon, the WSM based methodology will be referred to "crisp-logic based methodology" to compare with the "fuzzy-logic based methodology" proposed in the next chapter.

Figure 5.10: Top 10 route options according to our refined ranking method for restricted area on the right region.

Chapter 6

Fuzzy logic based methodology

Humans make decisions based on intuition and experience. Generally, these choices do not have any crisp-logic behind them. Similarly, a score given to a route option by human decision maker is also not fixed. Humans use quantitative methods to compare a set of options, but their decisions do not always follow a mathematical model. Moreover, different persons may score the same route differently. Even the same person may score a route route differently at various times or situations. Thus, to take into account the variance of scores a route can get for the same consideration, this research implements the proposed methodology using fuzzy-logic to mimic the human knowledge and approximate reasoning in the decision process. This will also allow the decision makers to dictate the importance of a decision consideration linguistically, something humans are more comfortable with, rather than define them in terms of integer values. The thesis expects the automated ranking process to be more user friendly by implementing it with fuzzy-logic because this will allow the users to define the decision criteria as high level linguistic rules.

In the previous chapters, the proposed methodology has been implemented using GIS, image processing, and WSM that incorporated crisp-logic to automatically rank route options. This automated ranking process for electrical transmission line has also been validated with industrial data and synthetic data. The implementation and validation proved the feasibility and potential of the proposed methodology in the electricity industry. To explore the four-step methodology with a different scoring method that tries to mimic the human reasoning, FIS is implemented instead of WSM in this chapter. The next section discusses the fuzzy based methodology. Section 6.2 implements and tests the proposed methodology using FIS and discusses the results. Section 6.4 validates the fuzzy-logic-based methodology with more route options and decision considerations and Section 6.5 compares it to the crisp-logic-based methodology and discusses their pros and cons.

6.1 Methodology procedures

The crisp-logic-based methodology proposed in Chapter 3 consists of four steps. To explore a fuzzy-logic-based model, the thesis uses the same four step methodology, except it implements Step 3 using FIS instead of WSM. Also, in fuzzy based model, Step 3 takes R as input as opposed to W. All the other steps remain the same, and so, they are not discussed in detail in this section. Please refer to Chapter 3 for details of Step 1, 2, and 4. This section only focuses on discussing the new fuzzy based Step 3 elaborately. Figure 6.1 shows the main steps, the inputs, and the expected output of the proposed methodology. Table 6.1 lists the symbols used in this work. The four steps of the methodology are explained in the following sub-sections.

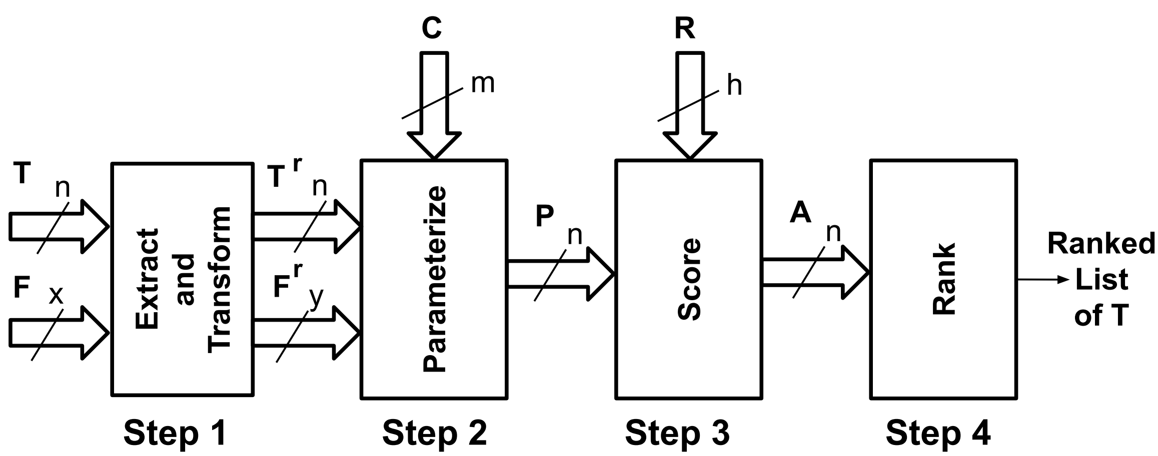


Figure 6.1: The fuzzy-logic-based methodology.

6.1.1 Feature extraction and transformation

The first step cleans data by extracting only the necessary features and converts the data to a common format by transforming them to raster format. All the data need to be in raster format as image processing techniques cannot be applied to maps in vector format.

Table 6.1: Nomenclature for fuzzy based methodology

Symbol	Description
F	a set of all natural and man-made features
x	number of features in F, $x = \|F\|$
F^e	a set of extracted features
y	number of features in F^e, $y = \|F^e\|$
f_k	an extracted feature belonging to F^e, $1 \leq k \leq y$
T	set of all route options
n	number of route options in T, $n = \|T\|$
t_i	a route option in the set T, $1 \leq i \leq n$
T^r	a set of transformed route options in raster format
t_i^r	a route option representing t_i in raster format
F^r	a set of extracted and transformed features in raster format
f_k^r	an extracted and transformed feature representing f_k in raster format
C	set of all decision considerations
m	number of decision considerations in set C, $m = \|C\|$
c_j	a decision consideration in the set C, $1 \leq j \leq m$
P	a set of parameters describing characteristics of route options
p_{ij}	a parameter describing a characteristic of t_i for decision consideration c_j
R	a set of fuzzy rules dictating higher level decision criteria
h	number of fuzzy rules in R, $h = \|R\|$
A	set of scores of all t_i
a_i	a score of t_i in set S
L	actual length on earth for each pixel
U	number of rows in a raster image
V	number of columns in a raster image

Formally, this step can be summarized as follows:

$$
\begin{aligned}
&F^e = \xi(F),\ \ F^e \subseteq F,\\
&F^r = \{f_k^r : f_k^r = \eta(f_k),\ \ f_k \in F^e, 1 \le k \le |F^e|\},\\
&T^r = \{t_i^r : t_i^r = \eta(t_i),\ \ t_i \in T, 1 \le i \le |T|\}
\end{aligned}
$$

where F^e is the set of extracted features, $\xi()$ is the extraction procedure, $\eta()$ is the transformation procedure, and $|X|$ is the cardinality of the set X. The sets F^e and T in raster format are represented by F^r and T^r, respectively.

6.1.2 Parameterization

This step finds the set of parameters P to describe and compare the route options based on C. It computes these parameters using T^r and F^r from the previous step. Each parameter $p_{ij} \in P$ describes the route option t_i^r for the decision consideration c_j. Thus:

$$
P = \{p_{ij} : p_{ij} = \lambda_j(t_i^r, c_j, F^r), 1 \le i \le |T^r|, 1 \le j \le |C|\}
$$

where $\lambda_j()$ is the parameterization procedure for the decision consideration c_j.

This step is repeated for all route options in T^r. To compare the route options $\{t_i^r\}_{i=1}^n$ for the specific decision consideration c_j, we simply compare the values of p_{ij} for $1 \le i \le n$.

6.1.3 Score Evaluation using FIS

In the third step, the methodology evaluates the route options and scores them. The thesis implemented Step 3 using WSM in Chapter 3. The crisp-logic-based methodology took P and W as inputs and produced A as output. It evaluated a cumulative score a_i for each route option in two sub-steps. In contrast, the fuzzy-logic-based methodology here evaluates a cumulative score a_i in three sub-steps using a FIS. It takes P and R as inputs and outputs A. Here, R is given as input instead of W and it is no longer a set of integers representing the weights of importance, rather it is a set of high level linguistic rules, fuzzy rules, that dictate the importance of the decision considerations in C. The fuzzy rules are given in the form of If-Then statements. Thus, the FIS takes the parameter values $\{p_{ij}\}_{j=1}^m$ describing a route option in t_i^r and a set of fuzzy rules R as inputs and outputs a score a_i between 0 to 100. This is repeated for all the route options in $\{t_i^r\}_{i=1}^n$.

The proposed method uses Mamdani Fuzzy Inference System Model [30]. The scoring method using FIS can be broken into three substeps: fuzzification, inferencing, and defuzzification. In fuzzification, it takes the crisp values of the parameters describing the route options and fuzzifies them based on the membership functions using the Mamdani implications. In inferencing substep, the resulting "clipped" membership functions are then superimposed using max operation to obtain the control inference. As the control inference is a fuzzy set, it needs to be defuzzified to obtain a crisp control value; in this case, the crisp control value is the route score. This defuzzification is performed using the centroid method. These steps can be expressed together as follows:

$$A = \{a_i : a_i = \phi_j(\{p_{ij}\}_{j=1}^{m}), 1 \leq i \leq |T^r|\}$$

where $\phi_j()$ is the FIS used in the scoring procedure. To compare all routes $\{t_i^r\}_{i=1}^{n}$, we simply compare the values in set A such that each a_i represents the score of t_i.

6.1.4 Ranking route options

In this last step, the methodology simply sorts T^r in descending order based on the values of the set A. The next section shows in detail how to apply the fuzzy-logic-based methodology.

6.2 Implementation using FIS

In this section, the fuzzy-logic-based methodology is used to rank a given set of route options for a transmission line in Saskatoon, SK, Canada that is shown in Figure 1.2. For this real scenario, the four input sets are defined as follows:

$$\begin{aligned}
&T = \{t_1, t_2, t_3, t_4, t_5\},\\
&F = \{f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”},\\
&\qquad f_5 = \text{“}residential\ area\text{”}, f_6 = \text{“}commercial\ area\text{”}\}, and\\
&C = \{c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}\},\\
&R = \{\ \textbf{If}\ (length\ is\ Short)\ \textbf{Then}\ (score\ is\ High)\\
&\qquad \textbf{If}\ (length\ is\ Medium)\ \textbf{Then}\ (score\ is\ High)\\
&\qquad \textbf{If}\ (length\ is\ Long)\ \textbf{Then}\ (score\ is\ Medium)\\
&\qquad \textbf{If}\ (ROWused\ is\ High)\ \textbf{Then}\ (score\ is\ High)\\
&\qquad \textbf{If}\ (ROWused\ is\ Medium)\ \textbf{Then}\ (score\ is\ Medium)\\
&\qquad \textbf{If}\ (ROWused\ is\ Low)\ \textbf{Then}\ (score\ is\ Low)\}
\end{aligned}$$

Here, the parameter, "length" is described using three linguistic variables: Short, Medium, and Long. "ROW used" and output score are described using another three linguistic variables: High, Medium, and Low.

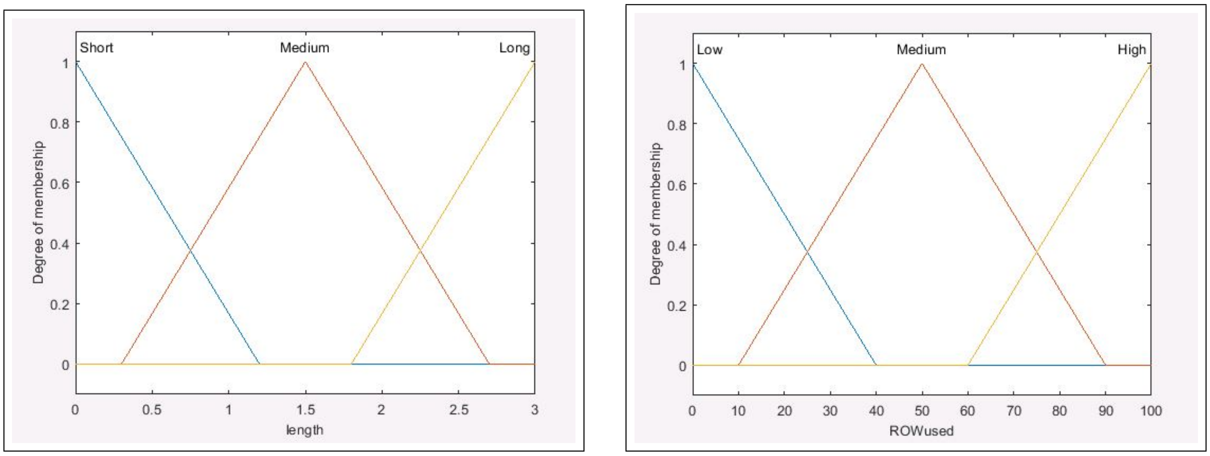
In the first step, the main objective is to obtain the set of extracted features F^r and the set of route options T^r in raster format from the input sets F and T. Given the two decision considerations in the set C, only features f_1 = "*roads*" and f_2 = "*railways*" have impacts on the route options. Therefore, $F^e = \{f_1, f_2\}$. The sets F^e and T are then converted to raster files as described in the methodology such that $F^r = \{f_2^r, f_2^r\}$ and $T^r = \{t_1^r, t_2^r, t_3^r, t_4^r, t_5^r\}$.

The second step *computes* the parameters needed to describe and differentiate the route options with respect to the decision considerations in C. The parameterization step is performed using MATLAB R2018a. It takes T^r, F^r, and C as inputs and outputs the set P. As there are two decision considerations in C, two parameters p_{i1} and p_{i2} are needed for every route option t_i^r, $1 \leq i \leq 5$. Here, p_{i1} represents the length of t_i and p_{i2} represents the ROW used by t_i. Thus, for this scenario, $P = \{p_{11}, p_{21}, p_{31}, p_{41}, p_{51}, p_{12}, p_{22}, p_{32}, p_{42}, p_{52}\}$. The parameters computed in Step 2 are shown in Table 6.2.

Table 6.2: The values for the parameters p_{i1} and p_{i2} from fuzzy-logic-based methodology.

Options	p_{i1} **(Km)**	p_{i2} **(%)**
t_1	2.1440	96.2687
t_2	2.5360	81.2303
t_3	2.1720	76.4273
t_4	2.8360	83.3568
t_5	2.7920	82.9513

The third step *computes* the set of scores A based on the parameter values in P and fuzzy rules in R. The fuzzy rules in R becomes the knowledge base for the FIS. To score a route option t_1 for the decision consideration c_1 and c_2, we give the parameter p_{11} and p_{12} as inputs to the FIS. The FIS uses the Mamdani model and scores t_1 based on the two inputs on a scale from 1 to 100 as a_1. This step is repeated for t_2, t_3, t_4, t_5 which produces their corresponding scores a_2, a_3, a_4, a_5. The set of scores A for this scenario generated by the fuzzy-logic-based methodology is shown in Table 6.3 under columns "a_i" with the label "automated". The membership functions used for the inputs "length" in km and "ROW used" in % are shown in Figure 6.2(a) and Figure 6.2(b), respectively.



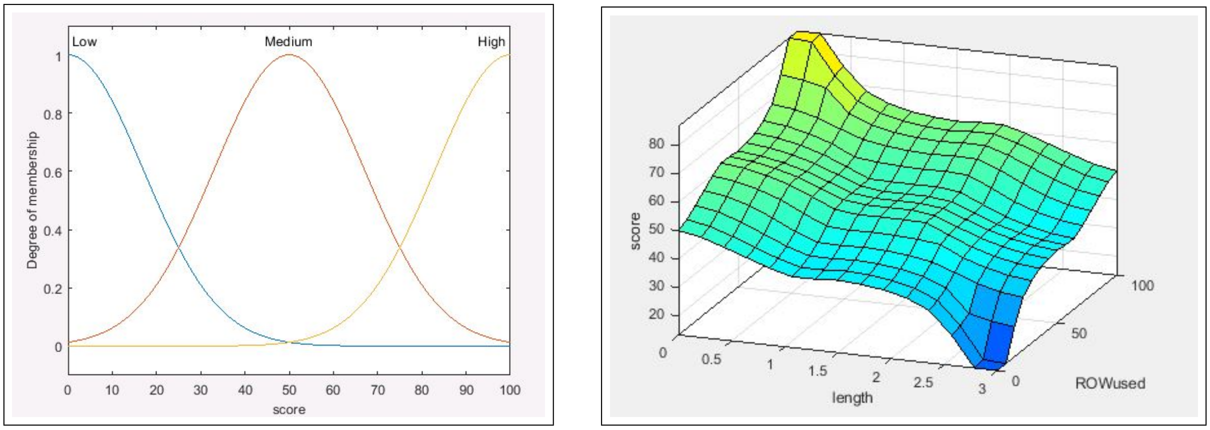
(a) (b)

Figure 6.2: The membership functions for inputs length (km) and ROW used (%) in the FIS are shown in (a) and (b).

The methodology uses triangular membership functions for the inputs. This is because, the length or the percentage can be represented as numbers and when they change, the degree of membership should also change at a constant rate. Thus, using linear function to represent the membership functions seems justified. For the output, we use Gaussian membership function. This is because different decision makers may score differently for the same set of inputs. This randomness generally follows a Gaussian distribution and so we used Gaussian membership function to mimic the human judgment of scoring. The output membership function is shown in Figure 6.3(a). The output surface for the FIS with respect to the two inputs is illustrated in Figure 6.3(b). Thus, the fuzzy descriptors for the input "length" of t_i are "short", "medium", and "long". For the input "ROW used" used by t_i and the output "score", the descriptors are "low", "medium", and "high".

Thus, the FIS in this example consists of two inputs with triangular membership function and an output with Gaussian membership function. It uses Mamdani model with six rules R. The system diagram is illustrated in Figure 6.4.

The final step ranks the route options based on the set of scores A. This step takes A, which is evaluated by the FIS, as an input and outputs the ranked list of T.



(a) (b)

Figure 6.3: The output membership function for score is shown in (a) and the fuzzy inference system output surface with respect to the inputs is shown in (b).

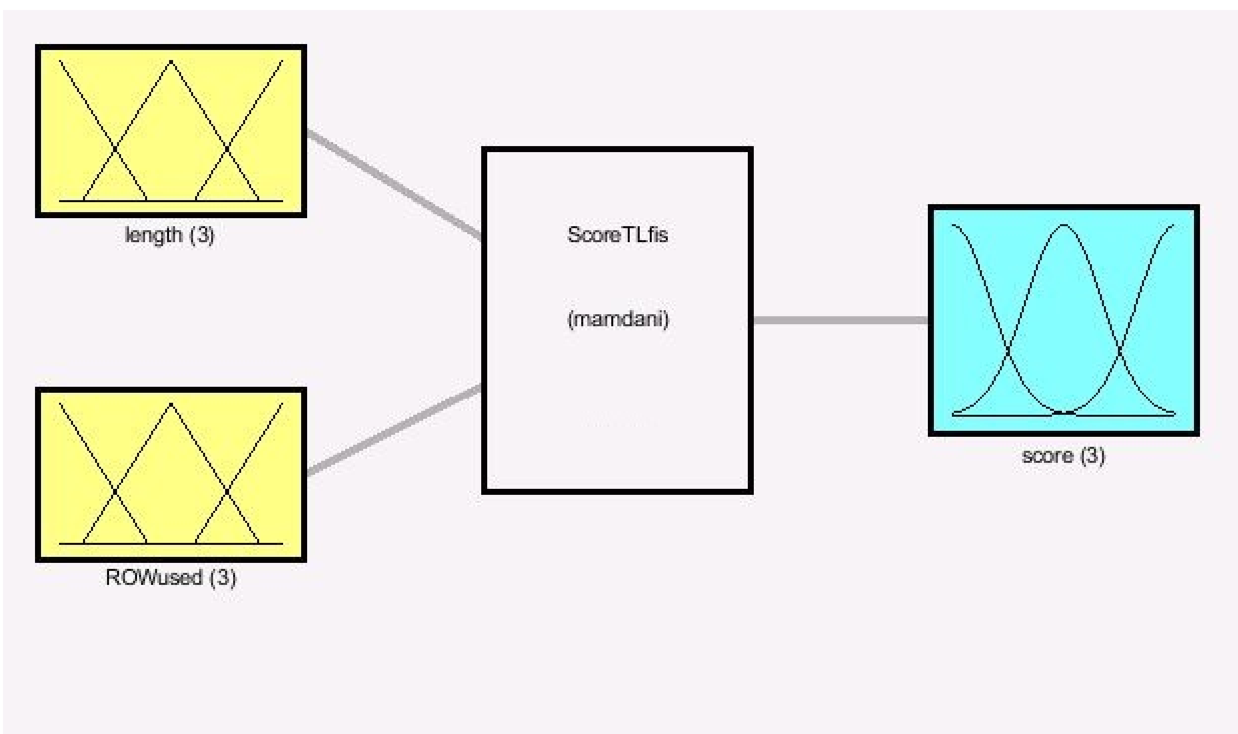


Figure 6.4: The Fuzzy Inference Model (FIS) used in the methodology.

6.3 Discussion

To evaluate the effectiveness of the fuzzy-logic-based methodology, this section compares the quality of the ranked route options with the manual method. The results obtained using the manual method are collected from a project in Saskatoon, SK, Canada, and they are shown in Table 6.3 with the label "manual". The results from the two methods are also illustrated in Figure 6.5.

Table 6.3: The scores generated by the fuzzy-logic-based methodology and manual method.

Options	a_i	
	automated method	**manual method**
t_1	59.2567	25
t_2	48.3850	19
t_3	61.2468	25
t_4	46.0929	13
t_5	45.8769	11

As shown in Figure 6.5, the fuzzy-logic-based methodology does agree with the manual method in ranking t_1 and t_3 as the top two optimal paths. It also accurately ranks t_2 as the third optimal option and t_4 higher than t_5 just like the manual method. However, it ranks t_3 higher than t_1 whereas the manual method ranked them equally. Moreover, we can also see the scores generated by the automated method is larger than the manual method. This is mainly due to the different methods used to score the route options. The automated method used the fuzzy inference system and scored on a scale of 0 to 100, whereas the manual method used human intuition and experience to score the route options. Other sources of error that could also have been introduced while calculating the length and ROW used. The average error in computing the length and ROW used are 7% and 4.2%, respectively. The errors are mainly due to the thinning process used for skeletonization, and scaling the path on the map. For example, in the process of reducing the thickness, the thinning procedure also reduces the length of the path slightly which is unwanted. In addition, when we multiply the number of pixels by L to find the length of the path scaled to what is on earth, we multiply the number of pixels with the distance represented by the width of the pixel. This introduces some errors when the path is diagonal because pixels are square in shape and the diagonal distance is longer than their sides. The accuracy of the fuzzy-logic-based methodology may also increase by increasing the resolution of the

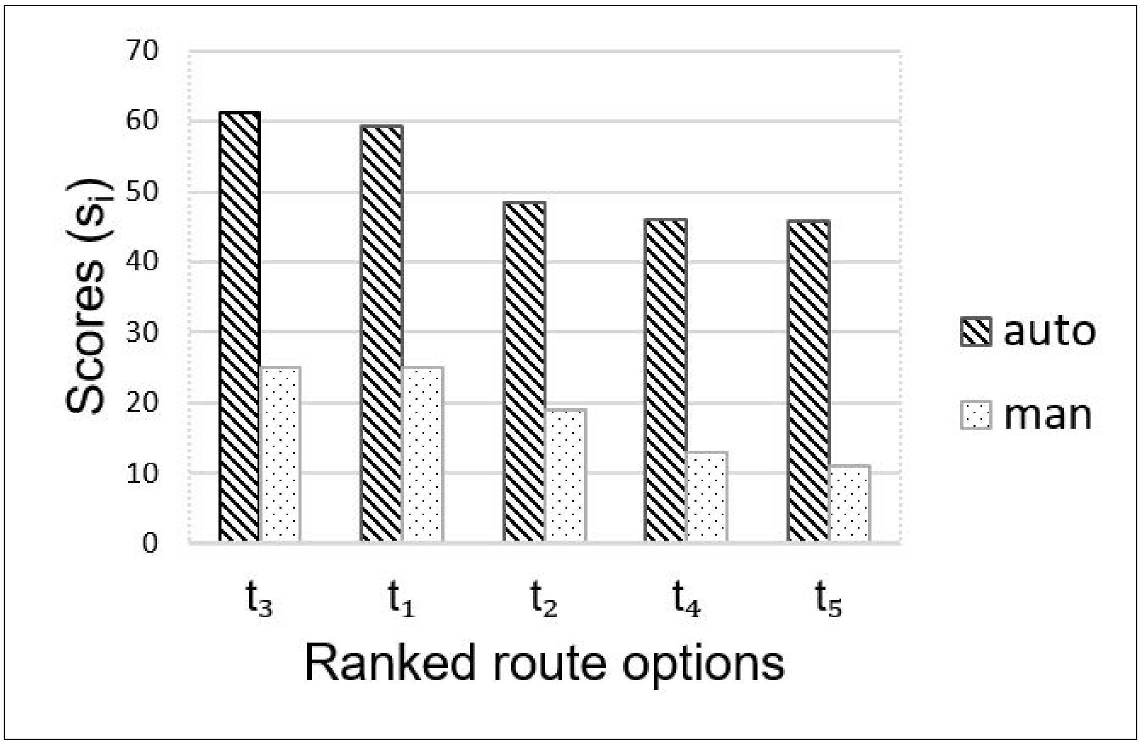


Figure 6.5: Scores generated by our methodology and the manual method.

images. However, it is to be noted that the pattern in scores is very similar to the manual method. As we are only interested in ranking the route options, the results have shown that our method produced almost the same ranking order as the manual method.

6.4 Validation

The previous section implemented and tested the fuzzy-logic-based methodology with real world data and compared the results obtained to industrial results. However, it was only tested for five route options and two decision considerations. To test the scalability and proper functionality of fuzzy-logic-based methodology, it is also tested with 155 routes and 6 decision considerations, which includes both manual and synthetic data. The synthetic data, routes and features, are generated using the *routePerturbation* and *featureGenerator* procedures described in Chapter 4. The inputs and outputs for this scenario are discussed in the following sub-sections.

6.4.1 Input

The proposed methodology using FIS is tested with 155 route options and 6 decision considerations. The input set is as follows:

$$
\begin{aligned}
T = \{&t_1, t_2, t_3, ..., t_{155}\},\\
F = \{&f_1 = \text{“}roads\text{”}, f_2 = \text{“}railways\text{”}, f_3 = \text{“}rivers\text{”}, f_4 = \text{“}parks\text{”}, f_5 = \text{“}residential\ area\text{”},\\
&f_6 = \text{“}commercial\ area\text{”}, f_7 = \text{“}restricted\ area\ 1\text{”}, f_8 = \text{“}restricted\ area\ 2\text{”},\\
&f_9 = \text{“}private\ area\text{”}, f_{10} = \text{“}forest\ area\text{”}\},\\
C = \{&c_1 = \text{“}cable\ length\text{”}, c_2 = ROW\ used\text{”}, c_3 = \text{“}restricted\ area\ 1\ used\text{”},\\
&c_4 = \text{“}commercial\ area\ used\text{”}, c_5 = \text{“}private\ area\ used\text{”},\\
&c_6 = \text{“}water\ area\ used\text{”}\}, and\\
R = \{\ &\textbf{If}\ (length\ is\ Short)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (length\ is\ Medium)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (length\ is\ Long)\ \textbf{Then}\ (score\ is\ Medium)\\
&\textbf{If}\ (ROWused\ is\ High)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (ROWused\ is\ Medium)\ \textbf{Then}\ (score\ is\ Medium)\\
&\textbf{If}\ (ROWused\ is\ Low)\ \textbf{Then}\ (score\ is\ Low)\\
&\textbf{If}\ (restricted\ area\ 1\ used\ is\ Low)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (restricted\ area\ 1\ used\ is\ Medium)\ \textbf{Then}\ (score\ is\ Medium)\\
&\textbf{If}\ (restricted\ area\ 1\ used\ is\ High)\ \textbf{Then}\ (score\ is\ Low)\\
&\textbf{If}\ (commercial\ area\ used\ is\ Low)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (commercial\ area\ used\ is\ Medium)\ \textbf{Then}\ (score\ is\ Medium)\\
&\textbf{If}\ (commercial\ area\ used\ is\ High)\ \textbf{Then}\ (score\ is\ Low)\\
&\textbf{If}\ (private\ area\ used\ is\ Low)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (private\ area\ used\ is\ Medium)\ \textbf{Then}\ (score\ is\ Medium)\\
&\textbf{If}\ (private\ area\ used\ is\ High)\ \textbf{Then}\ (score\ is\ Low)\\
&\textbf{If}\ (water\ area\ used\ is\ Low)\ \textbf{Then}\ (score\ is\ High)\\
&\textbf{If}\ (water\ area\ used\ is\ Medium)\ \textbf{Then}\ (score\ is\ Medium)\\
&\textbf{If}\ (water\ area\ used\ is\ High)\ \textbf{Then}\ (score\ is\ Low)\}
\end{aligned}
$$

Here, t_1 to t_5 are the manual routes, and t_6 to t_{155} are the randomly generated routes using the perturbation process. Features f_7 and f_8 are hypothetical restricted areas and are generated by the *featureGenerator* procedure. The membership functions for the input variables are shown in Figure **??**.

6.4.2 Output

For considerations c_1, we need the route options in T, where as for consideration c_2, we need features f_1 and f_2. Similarly, for considerations c_3, c_4, c_5, and c_6 we require features f_7, f_6, f_4, f_9, and f_3. Thus, in the first step, features extracted from F are $F^e = \{f_1, f_2, f_3, f_4, f_6, f_7, f_9\}$. The extracted features and transmission line route options are then transformed to raster format files that are used in the parameterization step. The methodology uses FIS in the third step to score the route options and sorts them in descending order according to their scores in the final step. The output generated by the methodology using FIS for the above input sets is shown in Figure 6.7. To check the proper functionality of the methodology, only the top 10 routes are illustrated in the Figure 6.8. In Figure 6.8, the red and yellow areas represent regions on the map with constraints and opportunities, respectively. The green lines are the top 10 routes options ranked by the automated ranking methodology proposed in this thesis.

6.4.3 Discussion

We can see from Figure 6.8 that the top ranked routes avoid the regions with higher constraints (areas in red) and utilize the corridors with opportunities (areas in yellow), but not as efficiently as the top ranked routes selected by the crisp-logic-based methodology in section 5.5. Some routes in Figure 6.8 pass through undesirable areas, such as restricted areas, water bodies, and commercial areas. Also, from Figure 6.7, we can see that the scores have less variance compared to the scores generated by the crisp-logic-based methodology. Thus, to analyse these results more critically, they are compared to the results from crisp-logic-based methodology in the next section.

6.5 Comparison

This section compares the outputs produced by the proposed methodology when implemented by two different scoring techniques: crisp and fuzzy logic-based techniques. Both the methodologies were tested for the same set of route options, features, and considerations $(T, F, \ and\ C)$. The crisp-logic-based methodology was given W and fuzzy-logic-based

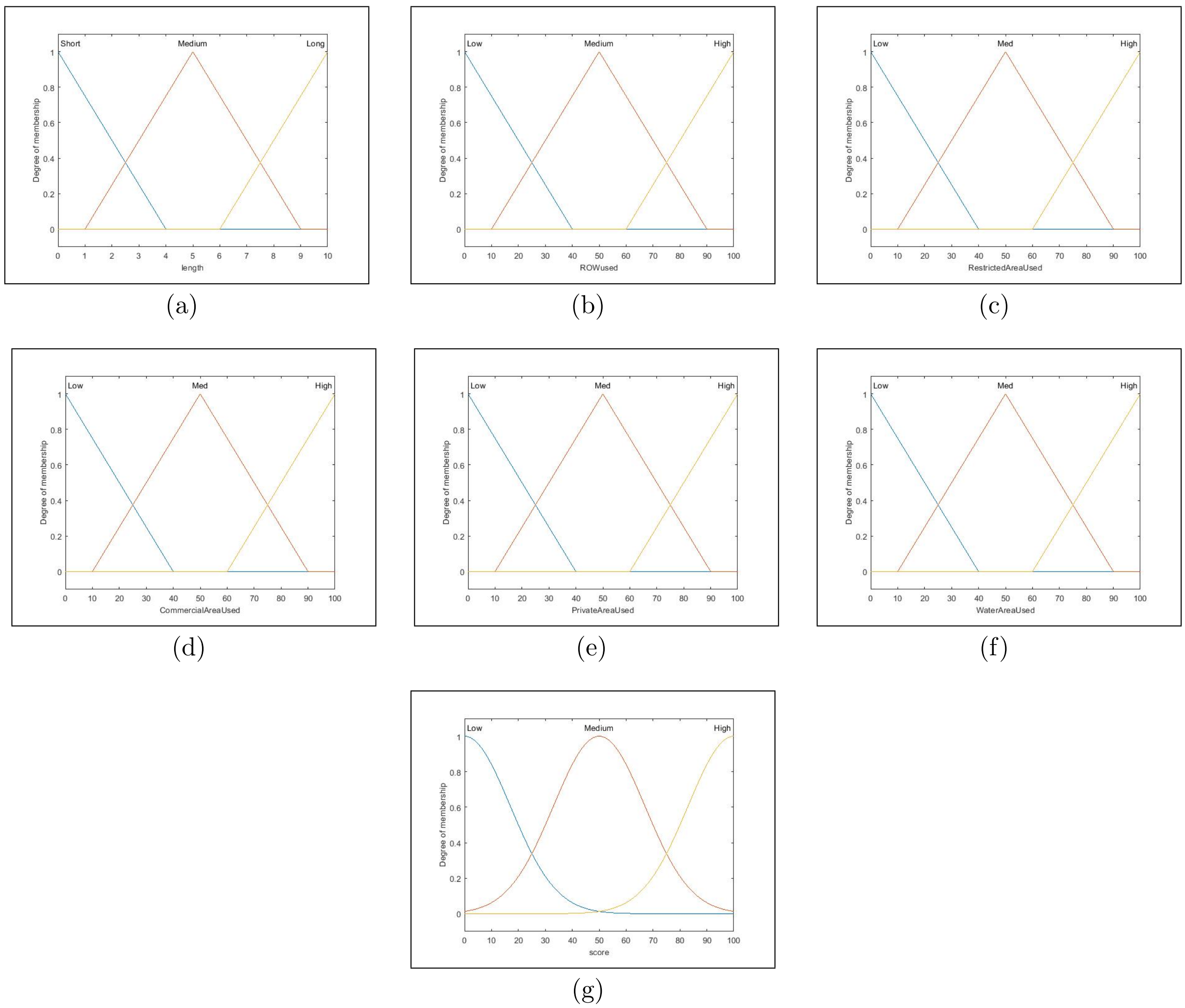


Figure 6.6: The membership functions for inputs, length (km), ROW used (%), Restricted area used (%), Restricted area used (%), commercial area used (%), private area used (%), and water area used (%) are shown in (a) through (f), respectively. The output membership function is shown in the part (g).

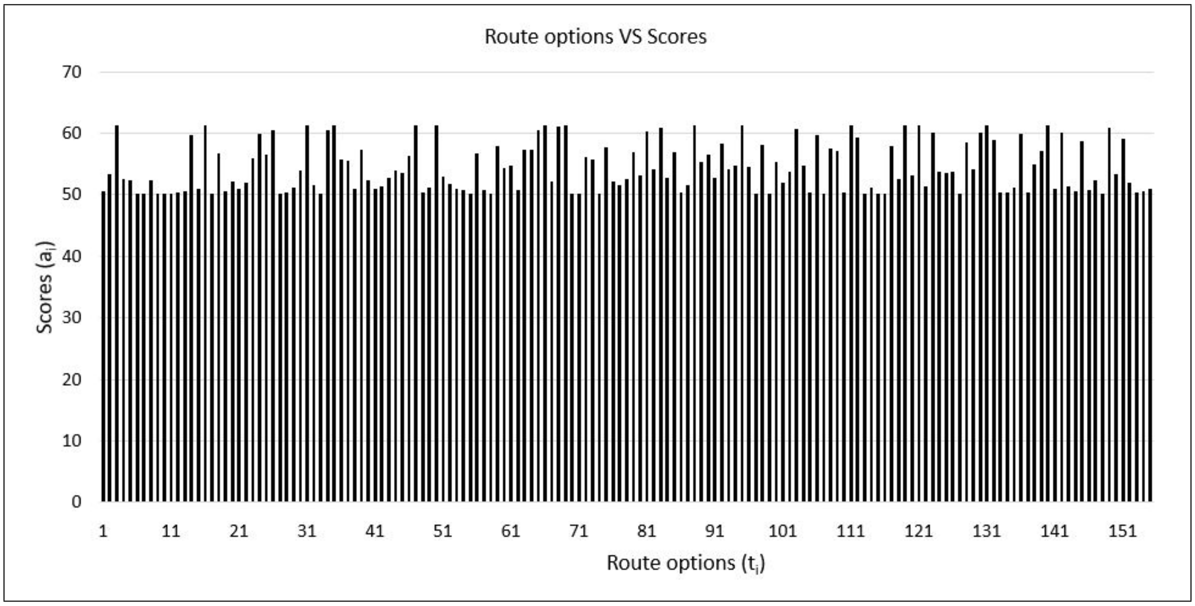


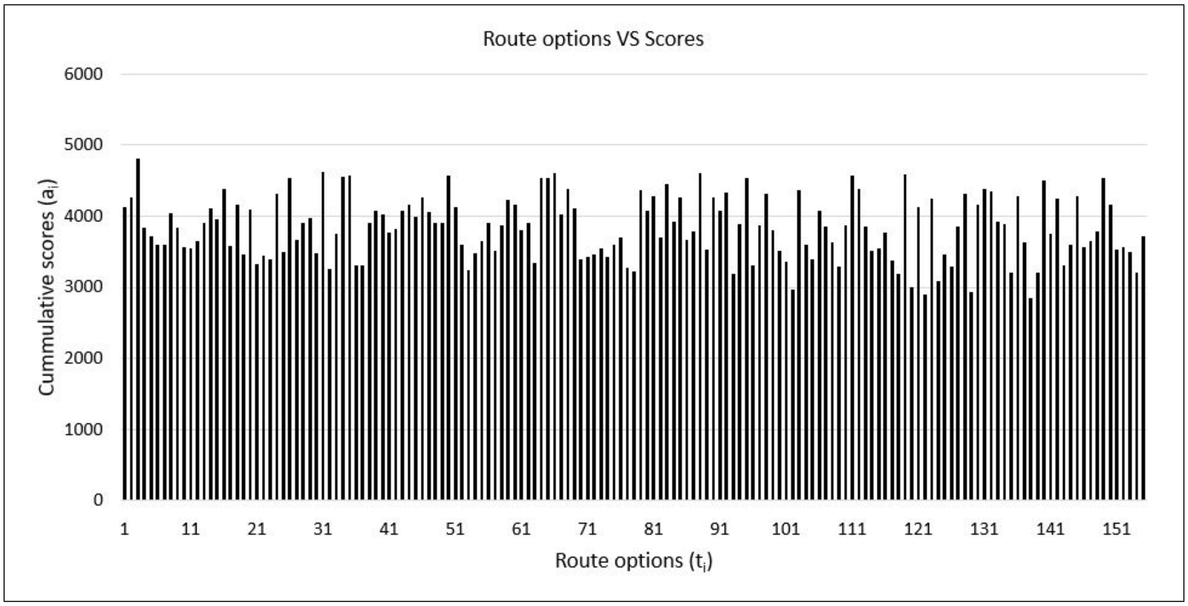
Figure 6.7: 155 route options and their scores according to the methodology using FIS when tested for 6 decision considerations.

methodology was given R as their fourth inputs. These inputs are discussed in detail in sections 5.5 and 6.4.

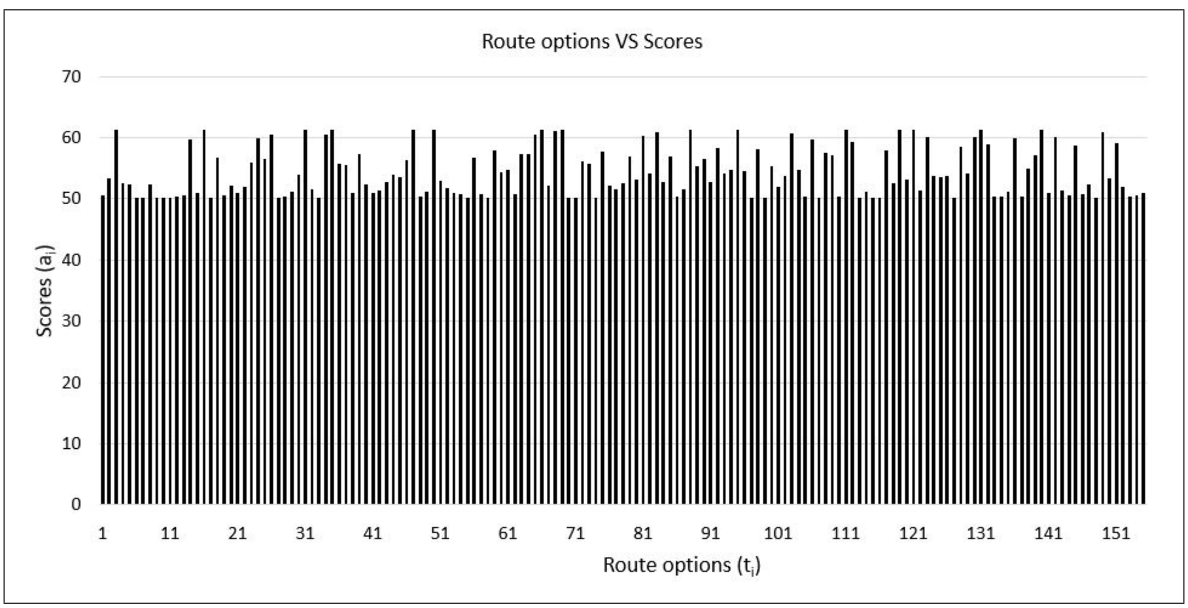
Figure 6.9 demonstrates the two sets of scores evaluated by crisp-logic-based methodology and fuzzy-logic-based methodology for the route options in the set T. Figure 6.9(a) shows that the crisp-logic-based methodology scored the route options in T with a minimum score of 2834 and a maximum score of 4794. The standard deviation of the scores is calculated to be 463.71 which makes the coefficient of variance equal to 11.4%. On the other hand, Figure 6.9(b) shows the scores by the fuzzy-logic-based methodology with a lowest score of 50.0022 and a highest score of 61.2204 The standard deviation of score is found to be 3.96 making the coefficient of variance equal to 7.29%. Thus, for the current parameters and settings, the crisp-logic-based methodology has slightly higher capability of differentiating between the route options. However, if we look at the Venn-diagram of the top 20 routes ranked by the two methods shown in Figure 6.10, we see that 16 out of 20 routes are common. In Figure 6.10, the top 20 route options according to the crisp-logic-based methodology are shown within the yellow circle labelled "WSMr" and the top 20 routes according to the fuzzy-logic-based methodology are shown inside the blue circle labelled "FIS". Here, the numbers inside the circles represent the subscripts i of the top ranked transmission line route options t_i and the number next to i in parenthesis shows the

Figure 6.8: Top 10 route options according to our fuzzy based ranking method for restricted area on the right region.

rank it received by its ranker. The routes that were among the top 20 in both the methods are shown in the green intersection of the Venn-diagram. The top 20 routes and the ranks they received by the two methods are listed in Table 6.4. Here, the ranks with an asterisk (*) show how a route chosen as top route by one method was ranked differently by the other. This difference can be minimized by fine-tuning the rules R in the fuzzy-logic-based methodology. The intersection in Figure 6.7 shows the outputs by the two methods, Crisp and fuzzy based, have a similarity of around 80%, implying that despite the differences and limitations, both methods produce similar outputs, and either method can be used in the industry in accordance to planners' preferences.



(a)



(b)

Figure 6.9: The scores of route options in T where parts (a) and (b) show scores generated by the methodology using WSM and FIS, respectively.

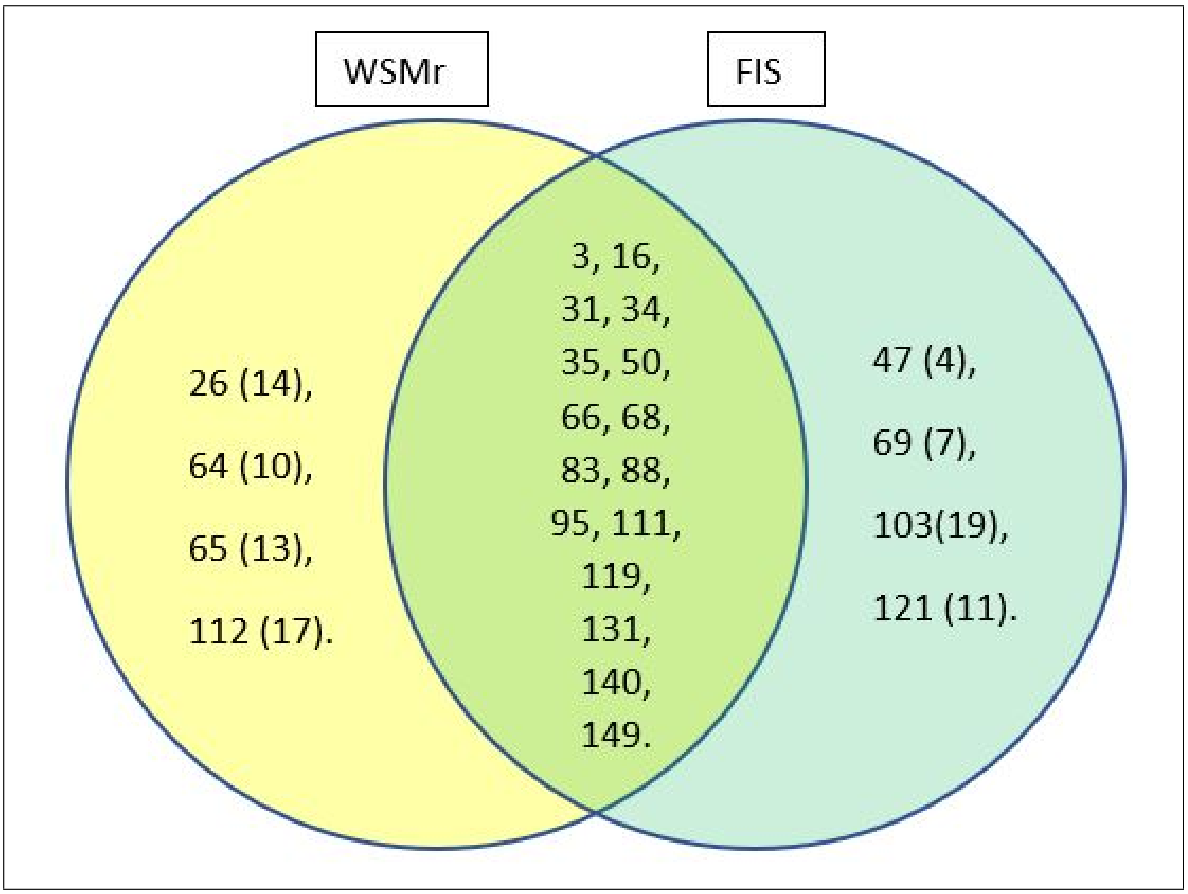


Figure 6.10: A Venn diagram showing the top 20 routes among 155 options according to the two methods: WSMr and FIS. The intersect area shows the routes that were ranked top by both methods.

Table 6.4: Top 20 routes ranked by the two methods.

	Route number	WSMr rank	FIS rank
Routes ranked top by both WSMr and FIS	3	1	1
	31	2	3
	88	3	8
	66	4	6
	119	5	10
	50	6	5
	35	7	13
	111	8	15
	34	9	20
	149	11	17
	95	12	9
	140	15	12
	83	16	18
	131	18	14
	16	19	2
	68	20	16
Routes ranked top by only WSMr	64	10	43*
	65	13	22*
	26	14	21*
	112	17	32*
Routes ranked top by only FIS	47	55*	4
	69	48*	7
	121	43*	11
	103	21*	19

Chapter 7

Conclusion

Ranking a set of route options is significantly important when selecting the optimal route for an electrical transmission line. This ranking is a complex process and it requires decision makers to consider many factors, making the traditional manual process time-consuming and erroneous. Different works have suggested the use of various decision considerations in selecting the optimal path. This thesis has classified the decision considerations that planners take into account during the route selection process into four categories: engineering, economic, social, and environmental. Then, it proposed a crisp-logic-based methodology to automate the process of ranking a set of feasible route options for a transmission line. The proposed methodology uses a GIS, simple image-processing techniques, and a WSM. The methodology was tested with a real scenario for a transmission line in Saskatoon, SK, Canada. The results showed similar patterns to industrial results reached based on human intuition and experience.

To validate the methodology further, the thesis generated synthetic data and tested it with various simulated scenarios. The research used perturbation and image processing techniques to generate random routes and hypothetical features. These synthetic routes and features were used to test the methodology with more route options and decision considerations, respectively. In the process of validation, it also improved the accuracy of the methodology by refining the WSM, which was then termed the WSMr. The methodology using WSMr successfully produced expected results when tested with one hundred and fifty five routes and taking six decision considerations into account. It proved the accuracy, functionality, and scalability of the proposed methodology, and its potential to be used in the electrical industry.

To make the automated ranking method mimic the human decision making process and be more user friendly, the thesis research also explored and implemented the proposed methodology using FIS instead of WSM. FIS allows users to define their preferences using linguistic variables. It performed with similar accuracy to the crisp-logic-based methodology and produced results that were around 80% similar to the other's. However, they are both unique approaches as FIS tries to imitate human knowledge and approximate reasoning, while, WSM is based on additive utility assumption. Despite their differences, both methods produced acceptable results and have the potential to be used in the electrical industry. Thus, the thesis presented two ways of automating the process of ranking routes for an electrical transmission line.

In the future, the work can be extended by incorporating more decision considerations, such as the number of bends in the route, estimated cost, and time needed to build a transmission line on that route. The proposed models can be improved by maximizing the number of common top routes and minimizing the difference between the routes' normalized scores. One of the limitations of this thesis is that the methodology was tested with one real world scenario due to unavailability of other real projects' data. Thus, in the future it should be tested with more real projects' data, in other regions, to confirm its use to the industry. Future works can also implement the methodology with other methods (e.g, using clustering techniques, neural network, etc.) to score the route options and compare them to the current models. Studying the impacts of raster file image resolution on result accuracy would also be worthwhile.

References

[1] Arcgis desktop help 9.3 - projection basics the gis professional needs to know. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Projection%20basics%20the%20GIS%20professional%20needs%20to%20know. (Accessed on 11/29/2019).

[2] Electricity customers — energy and the environment — us epa. https://www.epa.gov/energy/electricity-customers. (Accessed on 11/29/2019).

[3] Energy production & changing energy sources - our world in data. https://ourworldindata.org/energy-production-and-changing-energy-sources. (Accessed on 11/01/2019).

[4] Federal energy regulatory commission. https://www.ferc.gov/default.asp. (Accessed on 02/01/2019).

[5] Fuzzy logic - inference system - tutorialspoint. https://www.tutorialspoint.com/fuzzy_logic/fuzzy_logic_inference_system.htm. (Accessed on 12/19/2019).

[6] Fuzzy logic membership function. http://researchhubs.com/post/engineering/fuzzy-system/fuzzy-membership-function.html. (Accessed on 12/19/2019).

[7] A gentle introduction to gis. https://docs.qgis.org/3.4/en/docs/gentle_gis_introduction/. (Accessed on 11/29/2019).

[8] ghccrp_mcf_volume_1_121515.pdf. https://www.ct.gov/csc/lib/csc/pendingproceeds/docket_474/1_application/bulk/ghccrp_mcf_volume_1_121515.pdf. (Accessed on 11/29/2019).

[9] Gis file formats ˜ gis lounge. https://www.gislounge.com/how-to-tell-gis-file-formats/. (Accessed on 11/29/2019).

[10] Our renewable future. https://ourrenewablefuture.org/introduction/. (Accessed on 11/01/2019).

[11] Qer_appendixc_electricity.pdf. https://www.energy.gov/sites/prod/files/2015/09/f26/QER_AppendixC_Electricity.pdf. (Accessed on 11/28/2019).

[12] The structure of the alberta electricity market - energy and natural resources - canada. http://www.mondaq.com/canada/x/535890/Oil+Gas+Electricity/The+Structure+Of+The+Alberta+Electricity+Market. (Accessed on 11/28/2019).

[13] Types of energy - knowledge bank - solar schools. https://www.solarschools.net/knowledge-bank/energy/types. (Accessed on 11/15/2019).

[14] Types of gis data - gis (geographic information systems), geospatial data - research guides at syracuse university libraries. https://researchguides.library.syr.edu/c.php?g=258118&p=1723814. (Accessed on 11/29/2019).

[15] Types of gis data explored: Vector and raster ˜ gis lounge. https://www.gislounge.com/geodatabases-explored-vector-and-raster-data/. (Accessed on 11/29/2019).

[16] Vertical integration. https://www.investopedia.com/terms/v/verticalintegration.asp. (Accessed on 11/15/2019).

[17] What are map projections?—help — arcgis for desktop. http://desktop.arcgis.com/en/arcmap/10.3/guide-books/map-projections/what-are-map-projections.htm. (Accessed on 11/29/2019).

[18] What is a map projection? ˜ gis lounge. https://www.gislounge.com/map-projection/. (Accessed on 11/29/2019).

[19] About electricity — natural resources canada. https://www.nrcan.gc.ca/energy/electricity-infrastructure/about-electricity/7359, Jan 2019. (Accessed on 11/01/2019).

[20] Stefano Bagli, Davide Geneletti, and Francesco Orsi. Routeing of power lines through least-cost path analysis and multicriteria evaluation to minimise environmental impacts. *Environmental Impact Assessment Review*, 31(3):234–239, 2011.

[21] Lev M Bugayevskiy and John Snyder. *Map projections: A reference manual*. CRC Press, 2013.

[22] Jeff Christian and Lana Shipley. Electricity regulation in canada: overview — practical law canada. https://ca.practicallaw.thomsonreuters.com/5-632-4326?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1. (Accessed on 11/01/2019).

[23] Rodolfo Mendes de Lima, Reinis Osis, Anderson Rodrigo de Queiroz, and Afonso Henriques Moreira Santos. Least-cost path analysis and multi-criteria assessment for routing electricity transmission lines. *IET Generation, Transmission & Distribution*, 10(16):4222–4230, 2016.

[24] Semiye Demircan, Musa Aydin, and S Savas Durduran. Finding optimum route of electrical energy transmission line using multi-criteria with q-learning. *Expert Systems with Applications*, 38(4):3477–3482, 2011.

[25] Hasan Eroğlu and Musa Aydın. Genetic algorithm in electrical transmission lines path finding problems. In *2013 8th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 112–116. IEEE, 2013.

[26] Hasan Eroğlu and Musa Aydin. Solving power transmission line routing problem using improved genetic and artificial bee colony algorithms. *Electrical Engineering*, 100(3):2103–2116, 2018.

[27] Turan Gonen. *Electrical power transmission system engineering: analysis and design*. CRC press, 2015.

[28] RC Gonzalez, RE Woods, and SL Eddins. Digital image processing using matlab" gatesmark publishing, 2009.

[29] Sriram Kalaga and Prasad Yenumula. *Design of electrical transmission lines: structures and foundations*. CRC Press, 2016.

[30] FO Karray and C De Silva. Soft computing and intelligent systems design, 560 pp, 2004.

[31] Melita Kennedy, Steve Kopp, et al. *Understanding map projections*. Esri Redlands, CA, 2000.

[32] GISP Kurt Menke, Richard Smith Jr, Luigi Pirelli, GISP John Van Hoesen, et al. *Mastering QGIS*. Packt Publishing Ltd, 2016.

[33] Yongfu Li, Qing Yang, Wenxia Sima, Jiaqi Li, and Tao Yuan. Optimization of transmission-line route based on lightning incidence reported by the lightning location system. *IEEE Transactions on Power Delivery*, 28(3):1460–1468, 2013.

[34] A Luchmaya, B Dwolatzky, and AS Meyer. Using terrain information in an electrification planning tool. In *2001 IEEE/PES Transmission and Distribution Conference and Exposition. Developing New Perspectives (Cat. No. 01CH37294)*, volume 1, pages 456–460. IEEE, 2001.

[35] Claudio Monteiro, Ignacio J Ramírez-Rosado, Vladimiro Miranda, Pedro J Zorzano-Santamaría, Eduardo García-Garrido, and L Alfredo Fernández-Jiménez. Gis spatial analysis applied to electric line routing optimization. *IEEE transactions on Power Delivery*, 20(2):934–942, 2005.

[36] Pawel Pytlak and Petr Musilek. An intelligent weather-based system to support optimal routing of power transmission lines. In *2010 IEEE Electrical Power & Energy Conference*, pages 1–6. IEEE, 2010.

[37] Abd Manan Samad, Noorita Sahriman, Rosnani Anggong, Fazlina Ahmat Ruslan, and Ramli Adnan. A study of sabah electricity sdn. bhd.(sesb) best route transmission line using ahp. In *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA)*, pages 295–300. IEEE, 2016.

[38] S Ghandehari Shandiz, G Doluweera, WD Rosehart, L Behjat, and JA Bergerson. Investigation of different methods to generate power transmission line routes. *Electric Power Systems Research*, 165:110–119, 2018.

[39] Thomas Allen Short. *Electric power distribution handbook*. CRC press, 2018.

[40] Leonhard Späth and Anna Scolobig. Stakeholder empowerment through participatory planning practices: the case of electricity transmission lines in france and norway. *Energy research & social science*, 23:189–198, 2017.

[41] Zarko Sumic, SS Venkata, and Todd Pistorese. Automated underground residential distribution design. i. conceptual design. *IEEE Transactions on Power Delivery*, 8(2):637–643, 1993.

[42] Evangelos Triantaphyllou. Multi-criteria decision making methods. In *Multi-criteria decision making methods: A comparative study*, pages 5–21. Springer, 2000.

[43] Miguel Vega and Hector G Sarmiento. Image processing application maps optimal transmission routes. *IEEE computer applications in power*, 9(2):47–51, 1996.

[44] NA West, Barry Dwolatzky, and AS Meyer. Terrain based routing of distribution cables. *IEEE computer Applications in power*, 10(1):42–46, 1997.