GIS and Genetic Algorithm based Integrated Optimization for Rail Transit System Planning

Chro Haider Ahmed

Imperial College London

Centre for Transport Studies

Department of Civil and Environmental Engineering

A thesis submitted for the degree of Doctor of Philosophy

July 2016

Declaration of Originality

I hereby declare that I have personally carried out the entire work described in this thesis. Where source of information or the work of other have been used, they are fully cited and referenced and/or with appropriate acknowledgement given.

. . . .

Chro Haider Ahmed

Supervisors

. . . .

Professor Washington Yotto Ochieng

. . . .

Professor Michael Bell

Copyright Declaration

Acknowledgements

First and foremost, I would like to thank GOD, the author of knowledge and wisdom who gave me patience and power throughout the past four years and made this research possible.

I would like to express my sincere gratitude to my supervisor Professor Washington Ochieng for the continuous support of my Ph.D. study, for his motivation, valuable comments, and academic guidance through the rough road to finish my research. My sincere thanks and appreciation also goes to Dr Khalid Nur for his support, encouragement and academic guidance, without whose support it would not be possible to conduct this research. It is my privilege to thank Professor Michael Bell, for giving me the opportunity to join the family of Imperial College, for giving me the key to start my research journey and for making my dream come true.

I would like to thank the Higher Committee for Education in Iraq for their sponsorship of my study throughout the past four years. Also, I would like to express my appreciation to the university of Sulaimani, especially the Civil Engineering Department for giving me the study permission and for their support. In addition, I would like to express my heartfelt gratitude to my family who untiringly supported me spiritually and motivated me throughout my life.

I truly owe my deep gratitude to my beloved friends Mr. Karwan Nagada, his wife Mrs. Lana Talabani and their lovely kids for being my family in the UK, standing by me when times got rough and for their extra help and outpouring of support. My special thanks goes to the VIEW group, especially Mr Adil Mawalwi, Mr Nabaz Ghareeb and for my dearest friends and colleagues; Dr Yunis Alsheikhani and his beautiful family, Dr Serwan Zahawi and Dr Busyairah Syd Ali for sharing their greatly acknowledged support with me and who were always willing to help me. Finally, I wish to thank all concerned friends and relatives who continuously encouraged me.

Dedication

To my beloved parents

You have successfully made me the person I am becoming

And all my siblings

For supporting me all the way

Abstract

The planning of a rail transit system is a complex process involving the determination of station locations and the rail line alignments connecting the stations. There are many requirements and constraints to be considered in the planning process, with complex correlations and interactions, necessitating the application of optimization models in order to realize optimal (i.e. reliable and cost-effective) rail transit systems. Although various optimization models have been developed to address the rail transit system planning problem, they focus mainly on the planning of a single rail line and are therefore, not appropriate in the context of a multi-line rail network. In addition, these models largely neglect the complex interactions between station locations and associated rail lines by treating them in separate optimization processes. This thesis addresses these limitations in the current models by developing an optimal planning method for multiple lines, taking into account the relevant influencing factors, in a single integrated process using a geographic information system (GIS) and a genetic algorithm (GA). The new method considers local factors and the multiple planning requirements that arise from passengers, operators and the community, to simultaneously optimize the locations of stations and the associated line network linking them.

The new method consists of three main levels of analysis and decision-making. Level I identifies the requirements that must be accounted for in rail transit system planning. This involves the consideration of the passenger level of service, operator productivity and potential benefits for the community. The analysis and decision making process at level II translates these requirements into effective criteria that can be used to evaluate and compare alternative solutions. Level III formulates mathematical functions for these criteria, and incorporates them into a single planning platform within the context of an integrated optimization model to achieve a rail transit system that best fits the desired requirements identified at level I. This is undertaken in two main stages. Firstly, the development of a GIS based algorithm to screen the study area for a set of feasible station locations. Secondly, the use of a heuristic optimization algorithm, based on GA to identify an optimum set of station locations from the pool of feasible stations, and, together with the GIS system, to generate the line network connecting these stations. The optimization algorithm resolves the essential trade-off between an effective rail system that provides high service quality and benefits for both the passenger and the whole community, and an economically efficient system with acceptable capital and operational costs.

The proposed integrated optimization model is applied to a real world case study of the City of Leicester in the UK. The results show that it can generate optimal station locations and the related line network alignment that satisfy the various stakeholder requirements and constraints.

Published Papers

The following papers have been published in the course of this research:

Ahmed , C., Nur , K. and Ochieng, W. 2015, "A Novel Approach for Identifying the Locations of Potential Transit Rail Stations Using A GIS Based Model". In Proceeding of The First HCED Iraq Scholars Conference, London,UK.

Ahmed , C., Nur , K. and Ochieng, W. 2016, "Optimal Rail Transit System Planning Using Geographic Information System and Genetic Algorithm ". In Proceeding of 4th International Confrences on Advances In Civil, Structural and Mechanical Engineering (CSM), Birmingham,UK.

Table of Contents

List of Tables

Table of Figures

Chapter One – Introduction

Continuous growth of urban areas and the associated need for mobility have led to increasing transport problems, such as congestion, increased travel time and air pollution. The establishment of a rail transit system, such as surface trains, light rail, and metros can play a vital role in relieving all three of these problems. At the same time, it can also provide safer, more reliable and more convenient services for the population within major corridors and to important activity areas compared to other public surface transport systems. These potential benefits of rail transit systems have been identified in several studies. For example, Litman (2012) shows that per capita congestion delay, vehicle mileage and traffic fatality rates are considerably lower in cities with large rail transit systems than in comparable cities with no rail services. Gleave (2005) reports that, in the UK, established urban rail transit systems incur lower death and injury rates than other comparable surface transport modes.

These observations on the benefits of rail transit systems in alleviating transport problems have over the last few decades stimulated the expansion of existing rail transit systems and the construction of new ones in many cities around the world.

1.1 Research Background

A rail transit system has significant influences on the travel patterns, land use development, and economic, social and environmental characteristics across potential service areas. It involves massive investment in structures that are very difficult to relocate after construction. Therefore, rigorous planning and evaluation processes are critical to the acquisition of optimal rail transit systems. The planning of a rail transit system is a very complex process, which involves identification of station locations, networks of lines connecting stations, track geometry, bridges, tunnels and other system components. The network of lines and associated set of stations represents the principal infrastructure of a rail transit system and therefore, underpin the concept of operations involving the collection/distribution of passengers and their transportation over distance (Vuchic, 2005). These two interlinked components of the rail transit system therefore, represent crucial parts of the planning process.

The determination of the locations of stations and the rail lines that link them, entail complex decision making and evaluation processes. It involves the consideration of various local factors, including travel demand patterns, socio-economic growth, topography, land use, rights of way, existing roadway networks and soil conditions. Furthermore, the planning process must account for a number of constraints that arise from passenger, operator and environmental requirements. Therefore, there are many requirements, with complex correlations and interactions, to be considered, necessitating the application of optimization models in order to realise optimal (reliable and cost-effective) rail transit systems.

In traditional rail transit system planning, rigorous optimization methods that account for all the above factors are not applied. Oversimplification of the problem based on many assumptions is employed leading to the selection of a few candidate stations and corridors for further analysis before a final plan is chosen (Laporte et al., 2007). This approach has the obvious risk of not achieving an optimal rail transit system that considers all the relevant factors. To overcome the weaknesses of the traditional methods, research over the last few decades has formulated the planning of a rail transit system as an optimization problem and used operation research techniques to attempt to solve it. Despite their capabilities in allowing consideration of a higher number of candidate solutions than the traditional planning method, and performing trade-offs between multiple factors and constraints, such methods have a number of limitations, as summarized below.

- They mainly focus on a single rail line and, therefore, are not appropriate for the planning of networks with multiple lines.
- They largely neglect the complex interactions between rail lines and station locations by treating line alignment or station location in separate optimization processes.
- Crucial aspects of rail transit system cost effectiveness and efficiency, local conditions, and passenger and operator requirements, are only partially covered.

As a result, existing models for rail transit system planning are not generically optimal. Therefore, it is vital that new models are developed that satisfy the various passenger, operator and local requirements. Such a model would facilitate integrated rail transit system planning for multiple rail line systems. Incorporating the planning of multiple rail lines and associated station set locations into an integrated optimization process is also important in terms of enhancing the system performance and minimizing overall system costs. Lai and Schonfeld

(2016) showed that incorporating the planning of a single rail line and station location designations into an integrated optimization process reduced the total system costs, capital costs, and total operating costs and passenger cost by up to 58 % , 27%, and 12.5% respectively. Extending this to multi- line rail network should result in further benefits.

1.2 Research Aim and Objectives

Given the background to and limitations of the state-of-the-art in the rail system planning process identified above, the aim of this research is to develop an optimal planning method that treats the rail network system and its influencing factors in a single integrated process. Five objectives have been formulated to achieve this aim.

1- Identify the requirements for rail transit system planning with respect to the passenger level of service, operator productivity, and potential benefits to the community, each of which have a significant influence on both the location and configuration of the rail transit system.
2- Disaggregate the identified requirement sets of the three interrelated parties: passengers, operators and the community into group sets according to their interactions with the two main components of rail transit system planning (station locations and the alignment of the line network connecting the stations).
3- Quantify and formulate various associated station and rail line network planning requirements as an optimization problem to achieve an efficient and effective rail transit system.
4- Develop an effective method to seek the best solution for the rail transit system planning problem with respect to the formulated set of the identified requirements and constraints.
5- Conduct a real world case study to examine the effectiveness of the proposed method and to confirm its validity.

1.3 Thesis Structure

The thesis structure is outlined below.

Chapter 1: Introduction

This chapter introduces the background to rail transit system planning, defines the problem and state the aim and objectives of the research.

Chapter 2: Trends and Developed Models for Rail Transit System Planning

This chapter reviews the state-of-the-art in the development of models for rail transit system planning. The review focuses on the representation of the planning procedures used in generating various alternative solutions, the selection of criteria to evaluate alternative solutions, the incorporation of various environmental and geographical constraints and, the demonstration of various search methods to obtain the objective solution to the problem. The reviews aim to find the gaps/limitations in the state-of-the-art and thus identify those aspects that need further improvement in the planning process. Based on the highlighted gaps/limitations of the existing models this chapter also presents the novel aspects of the problems addressed in this thesis key contribution of the research.

Chapter 3: A New Framework for Rail Transit System Planning

Based on the limitations identified in the existing models and on the objectives of this research, this chapter develops a framework for rail transit system planning. The framework designed to enable rail transit planners to determine station locations and rail line networks connecting the stations based on trade-offs between various rail transit system requirements and constraints. After specifying the framework, this chapter addresses the requirements function or module. It identifies the requirements of the various rail transit system stakeholders (passengers, operators and the community) and the factors that influence them. The requirements are then formulated within the context of an integrated optimization process that can determine station locations and associated line networks simultaneously.

Chapter 4: Determination of Potential Rail Transit Station Locations

This chapter develops a GIS based algorithm to determine potential station locations. The algorithm measures the feasibility of station sites and identifies the candidate pool of potential station locations with a comprehensive consideration of the various requirements (as identified in chapter 3), based on systematic evaluation and comparative analysis. This chapter also conducts a sensitivity analysis to demonstrate the impact of the different requirements on the location and number of generated candidate solutions.

Chapter 5: Optimization of Station Locations and Associated Line Network

This chapter details the development of a heuristic optimization algorithm to simultaneously select the best set of the potential stations from the solution set identified in chapter 4 and generate an optimal line network to connect the selected stations. The optimization algorithm, is based on genetic algorithm (GA), supported by a geographic information system (GIS) database. It accounts for complex correlations and interactions of the various requirements identified in chapter 3, and makes trade-offs between them to achieve an optimal (reliable and cost effective) system. This chapter also elaborates how the optimization algorithm is designed to consider various local factors such as travel demand pattern, land use pattern, topography and soil conditions as well as outlining the key steps of the optimization process.

Chapter 6: Case Study and Sensitivity Analysis

This chapter explores a real-world planning scenario using the City of Leicester as the study area to examine the effectiveness of the proposed method. Extensive numerical analyses are undertaken in this chapter to: examine the importance of screening the study area for feasible station locations prior to the optimization process; demonstrate the advantage of simultaneous optimization of systems with multiple rail lines over individual optimization; tune, verify and compare different parameters of the genetic algorithm to find reasonable parameter values for solving the rail transit system planning problem; as well as evaluating the goodness of the solution obtained by the proposed model.

Chapter 7: Conclusions and Future Work

The main contribution of this thesis in relation to the stated aims and objectives are summarised in this chapter. Potential topics for future research in this field are presented also.

Chapter 2 - Current Models for Rail Transit System Planning

This chapter reviews the relevant literature on rail transit system planning. It formulates the rail transit system planning problem and reviews in detail the current techniques for solving the problem. It is organized into four main sections. Section 2.1 describes the formulation of the problem, elaborating the major sets of objective functions and constraints that are associated with rail transit system planning. Section 2.2 investigates various search algorithms and artificial methods that can be used to generate solutions to the planning problem. Section 2.3 reviews the existing approaches to solve the two interlinked aspects of rail transit system planning: determination of station locations and network of lines connecting the stations. Basically, these models represent a structured representation of the set of decision variables, objective functions, and constraints at a fairly aggregated level that can be used as a tool or strategy for decision-making within the context of the problem solving process. The chapter is concluded in section 2.4 by a summary of the findings from the literature review and the research questions addressed in this thesis.

2.1 Formulation of the Problem

The basic rail transit planning problem seeks to identify the best possible locations for stations and also to determine the best possible network of lines connecting these stations. This search takes into account various local conditions (e.g. travel demand patterns, land use patterns, rights of way, and existing road networks) while seeking to satisfy a number of complex and interrelated requirements. These include minimum travel time, adequate travel speed, minimum investment cost, and maximum population coverage. Some of these requirements partially overlap while others are in conflict. Partially overlapping requirements include minimum passenger travel time and maximum operating speed. Conflicting requirements, meanwhile, include minimum investment cost and maximum population coverage. Therefore, it is important to formulate the rail transit system planning process as an optimization problem in order to allow trade-offs between these conflicting requirements to achieve a reliable and cost-effective rail transit system.

Various optimization models have been proposed to resolve the essential trade-offs among the conflicting requirements. Basically, to represent the problem within an optimization context, it is essential to formulate its decision variables into a set of constraints and objective functions. Accordingly, the following sub-sections present how rail transit system planning is represented within the context of an optimization problem in the literature. Section 2.1.1 discusses in detail the objective functions that are commonly employed in the determination of rail transit station locations and the associated line network. This section also elaborates the parameters associated with the objective functions. Section 2.1.3 discusses the main constraints that are used in rail transit system planning.

2.1.1 Objective Functions Associated with Rail Transit System Planning

In general, four main objective functions are used to solve the rail transit system planning problem: passenger costs, operational costs, construction costs and the population or traffic coverage, as depicted in table 2.1. A number of studies have cast the problem within the context of multi-objective optimization (Laporte et al., 2011, Gutiérrez-Jarpa et al., 2013), while others have formulated it within the context of a single-objective optimization (Dufourd et al., 1996, Laporte et al., 2005, Lai and Schonfeld, 2010). Single-objective optimization entails planning the system to provide either (1) minimum passenger and/or operator costs subject to a budget constraint or (2) maximum population or traffic coverage subject to a budget constraint. The following sub-sections define the components of each of these objective functions in detail.

Table 2.1: Rail Transit System Objective Functions

Objective Functions	**Decision Variables**
Passenger Costs	Access time, Waiting time, In-train time
Operational Costs	Train/station/lines operation and maintenance
Construction Costs	Rights-of-way, Earth works, Structures
Population/ Traffic Coverage	Number of people /trips covered by the rail system

2.1.1.1 Passenger Costs

Passenger costs are those items covering passenger travel time from the origin to the desired destination, as well as the associated passenger cost items for utilizing the rail transit system service and access to/from stations, if passengers elect to drive to/from the access station. Passenger cost items are affected by different features of the rail transit system, such as the number, integrity and independency of network lines, as well as number and location of stations.

The travel time costs are those items which cover passenger access time to/from stations, waiting time at the station, and travel time onboard. Travel time often represents the dominant component of "disutility" to the passenger of a rail transit system. Hence, its minimization is one of the crucial objectives of rail transit planning. Passenger travel time on a rail transit line consists of three main elements (Chien and Schonfeld, 1997, Van Nes and Bovy, 2000):

1- **Access time:** is defined as the travel time from origin to the rail transit station and from the transit rail station to the destination. It depends on the distribution of the origins and destinations of passengers in relation to stations and routes, and the location and number of stations. Generally, four main travel modes can be used to access the rail transit system (Lai, 2012). They are summarized below and explained in detail in chapter three:

   i. *Walking:* is the common mode used to access rail stations in central cities. When walking is the only mode that passengers use to access the system, the access time component consists only of the travel time to/from station.

   ii. *Bike-and-Ride:* has emerged as a commonly used mode of accessing transit rail systems over the last few decades in many developed countries. In this case, the cycling time to station should be considered as the access time.

   iii. *Bus:* is another commonly used option for rail transit system access. When this mode is used, the access time consists of the waiting time at a bus stop plus the riding time in a bus to arrive at a rail transit station.

   iv. *Park-and-Ride:* is the most popular mode of accessing rail transit systems in low density suburban areas. In order to avoid the heavily congested surface corridors, passengers drive to stations through uncongested local roads and then use rail transit. The driving time to stations should be considered as a part of the access time. Furthermore, the cost of fuel and emissions associated with the access travel and the parking costs at a station if the parking service facility is operated by a private sector should be considered also when computing the cost of the access time.

2- **Waiting time at a station:** This depends on the frequency of trains, reliability of the train schedules and passenger arrival patterns at the stations. Usually for short headways (i.e. train frequencies), passengers do not follow schedules, and their average waiting time at the station is assumed to be half of the train headway (Chien and Schonfeld, 1997). On the other hand, in the case of long headways (usually >6 min) passengers start to use train schedules and regulate their arrivals in order to minimize their average waiting times. With increasing headway times, this becomes the norm and waiting times approach constant values (Bowman and Turnquist, 1981).

3- **Travel time in the train:** This depends on the distance between the boarding and alighting stations, geometric characteristics of the rail lines, station density, and train operating speeds. In addition, the way in which trains operate greatly influences this travel time element (Vuchic, 1981). For example, if rail transit network corridors are only partially separated from other surface transport system corridors, passengers may experience higher travel times compared to trains from networks with fully separated corridors. This is due to interactions with other surface traffic, such as potential intersection delays at-grade crossing.

2.1.1.2 Operational Costs

Operational costs cover the items associated with the operation and maintenance of trains, tracks, electric traction installations, stations, terminals, energy supply, signalling systems and traffic management and safety systems. The costs are incurred throughout the life span of the rail transit system. These cost items depend on a variety of factors, such as system size and form, including the density of stations, length, number, integrity and connectivity of line network connecting stations, as well as demand and various local conditions (e.g. rights-of-way and topography). Considering station density, for example, shorter inter-station intervals result in lower operating train speeds and higher operating and maintenance costs due to frequent train stops. In contrast, longer interstation intervals result in higher operating train speeds, thus in effect lowering the operating and maintenance costs. In general, therefore, the operational and maintenance costs increase with station density. The cost of maintaining tracks and electric traction installations depend on the length of the line network and the number of trains running during operations. Similarly, the operational costs of trains, which include the costs of labour and energy, largely depend on the number of trains running on each particular line during operations, which is indirectly determined by demand (de Rus, 2012).

Often, the operational cost of a rail transit system is measured in unit cost per passenger – kilometre (de Rus, 2012, TFL, 2014, ORR, 2012). This measure of how much, on average, it costs a train operating company to move a passenger over one kilometre provides a useful "headline" assessment for transport agencies and passengers of the cost of delivering a particular service. It is also important to note that once a service is already running, carrying additional passengers is not generally a major driver of operational costs, but is the primary revenue generator (ORR, 2012). For example, the operating costs per passenger-kilometre (pkm) on the London rail network were reduced by 26% between 2008/09 and 2013/2014 due to increased passenger numbers (TFL, 2014). However, it should be noted that a significant increase in passenger demand could result in the need for extra trains or the provision of additional services, and this could lead to significant increases in the consequent operational costs.

2.1.1.3 Construction Costs

These refer to the construction costs that are directly incurred by government or rail transit agencies. Typically, the construction of rail transit systems includes provisions for ensuring the right-of-way of the line, performing earthworks, fastening rails, building structures (e.g. bridges, tunnels, stations) and other such items. These construction items vary significantly with network size and form including the number of stations, types of stations, number and length of lines, as well as various local conditions, such as rights-of-way, topography and soil conditions. In conjunction with network size, construction costs can be broken down into three main groups with respect to rail line lengths (Hay, 1982).

- **Variable construction costs:** include certain cost items that vary directly with rail line length, such as rail tracks, ties, ballast, fences and guardrails. In general, these cost items account for about 20% of the construction costs.
- **Semi-variable construction costs:** include items that vary not only with length but also with various local conditions, such as travel demand, topography, geology, land use and land

values. Examples in this group include the number of stations, rights-of-way, bridges and tunnels.

- **Non-variable construction costs:** include cost items that have no relation to line lengths. Examples include major shops and office buildings within terminal facilities.

Among the various local conditions that have considerable influence on the construction costs in addition to rail transit system performance are rights-of-way, which include land acquisition, property damage and compensation. According to its interaction with both the investment in and operational efficiency of rail transit systems, the right-of-way element has three categories (Vuchic, 1981):

- **Right-of-way Category A:** represents transit corridors that are fully separated and physically protected from other surface traffic. These include tunnels, aerial (elevated) structures or fully protected at-grade tracks. Transit systems within this category involve substantial investment but also provide considerably higher performance in terms of speed, reliability, capacity, riding comfort and safety.
- **Right-of-way Category B:** represents transit corridors that are partially separated from other surface traffic. Typically this category includes the rail tracks on street median that are longitudinally separated, but with at-grade crossings for vehicles and pedestrians. Since this category is only partially separated, it involves considerably lower investment costs than category A. However, the transit system performance in this case is lower than that of category A due to interaction with other traffic.
- **Right-of-way Category C:** represents transit corridors in which trains and other surface traffic or pedestrians share operations. Usually this right-of-way category requires low to moderate investment. Here, however, the system performance is profoundly affected by traffic conditions along its line.

2.1.1.4 Population/Traffic Coverage

Maximizing the population within the system coverage (i.e., number of people living close to rail stations) or covered trips (i.e., actual number of people using the rail system) is one of the most widely used objective functions in planning rail transit systems. The degree to which the system contributes to increasing population coverage or transit trips is a key determinant in the ability of a rail transit system to alleviate traffic congestion, improve mobility, and reduce energy consumption and emissions. It is notable that some rail transit system planning studies incorporate maximum population coverage into the optimization framework (Dufourd et al., 1996, Bruno et al., 2002), while some others embed maximum trip coverage (Laporte et al., 2005, Gutiérrez-Jarpa et al., 2013). In general, the existing rail transit planning models that are based on maximum trip coverage apply the conventional four-step travel demand forecasting models, while those that are based on maximum population coverage employ the following two simple approaches to estimate the number of people covered by the system..

1- Alignment catchment area: consists of drawing embedded corridors around the alignment with an assigned coverage score to each corridor. It assumes that coverage gradually decreases as the access distance to the alignment increases and subsequently sets to zero. This is essentially the approach taken by Chapleau et al. (1987) and Wirasinghe and Vandebona. (1987). The major drawback of this approach is that it neglects the fact that people who live near a rail transit line but far from a station are less likely to use the system (Laporte et al., 2005).
2- Station catchment area: consists of drawing concentric geometric shapes with decreasing attraction factors around each station. It assumes that coverage gradually decreases as the access distance to the station increases and subsequently approaches zero. This approach is used by Dufourd et al. (1996) and Bruno et al. (2002). It offers a more realistic estimation of the covered population than the alignment catchment model, although it is based on census data without considering the socioeconomic, demographic and travel characteristics of the population.

From the rail transit system planning perspective, maximizing population coverage rather than trip coverage is sensible. This is to the extent that people tend to relocate over time to satisfy their travel requirements once the system begins operation, regardless of what the configuration of the system is, and subsequently this results in a change in the initial travel pattern. Nevertheless, capturing maximum trip coverage as opposed to population coverage is in fact more appropriate and realistic, particularly when the optimal layout of the line network needs to be determined. This is because the population coverage approach, as well as ignoring the socioeconomic, demographic and travel characteristics of the population, does not consider station-to-station usage. For example, a person seeking to travel in an East-West direction is unlikely to be attracted to a station located on a

North-South alignment, assuming that this alignment is not a part of a larger interconnected network (Laporte et al., 2005).

Maximizing population coverage, therefore, does not adequately capture the primary objective of a rail transit system which is to improve mobility (Laporte et al., 2000, Laporte et al., 2005, Gutiérrez-Jarpa et al., 2013). As a result, recent rail transit system planning models have largely used trip coverage as opposed to population coverage (Lai and Schonfeld, 2016, Repolho et al., 2013), even though this requires more substantial data gathering and computational effort. It should be noted that the accuracy with which the number of passengers that will use a proposed rail transit system can be predicted is vital for accurate passenger and operator cost calculations. Uncertainty in the passenger demand forecasts can have profound effects on the local investment, operator efficiency and passenger service quality. For example, the Skytrain project in Bangkok (Thailand) was greatly over-dimensioned due to a 2.5% overestimation of passenger forecasts. As a result, very large terminals where constructed at great cost, providing long platforms and train numbers. Once the system began operating, a large number of trains were unused due to a lack of demand and, consequently, the concession company encountered financial difficulties due to very low revenues (Flyvbjerg et al., 2004). To avoid such risks, demand forecasts must sufficiently capture traveller behaviour, land use patterns and transport infrastructure of the study area.

2.1.2 Constraints Associated with Rail Transit System Planning

A number of constraints have to be considered in rail transit system planning in order to achieve high efficiency in terms of both operations and functionality. These constraints can be categorized into two main groups; (i) design constraints and, (ii) environmental and geographical constraints. The former are related to operational requirements, while the latter vary with various local conditions, such as geography, land use patterns, geology and preferences of the community.

2.1.2.1 Design Constraints of Rail Transit Systems

The design constraints of rail transit systems can be divided into two main parts according to their relationship to the line network configurations and associated station sets;

1- Design constraints of the rail transit line network: these include the spacing between lines, length of lines, and connections between lines, as well as geometry requirement of line alignments (such as, minimum horizontal and vertical curve radius, and minimum and maximum gradient).
2- Design constraints of rail stations: these include the number of stations along each line of the system and the spacing between stations along each line of the system.

Several studies have incorporated these design constraints in rail line alignment and station location optimization processes due to their significant influence on the overall system performance and cost. For example, in order to restrain construction costs, Laporte et al. (2005) used a line length constraint in choosing stations among a set of candidate locations for a rail transit line. To guarantee an acceptable operation speed, Samanta and Jha (2011) integrated constraints in respect to the spacing and number of stations when optimizing station locations for a rail transit alignment. The incorporation of these constraints with the optimization process is useful for controlling the operational requirements and achieving an economically viable rail transit system solution.

2.1.2.2 Environmental and Geographical Constraints

Environmental and geographical constraints play an important role in the decision-making process for rail transit system planning, and therefore, needs to be accounted for in the optimization process. These constraints can be categorized as (Kang, 2008):
1- Environmentally sensitive areas: these include wetlands, floodplains, forests, historic and archaeological areas that should be avoided. The planning and construction of rail transit lines or stations across these regions are generally expected to encounter difficulties.
2- User preference areas: these include control areas or fixed points that planners consider need to be serviced by the rail transit system. These may include major activity centres, shopping centres, and freeway interchanges.

Chapter three provides further details on the above objective functions and constraints and their importance in planning rail transit systems within the broad context of passenger, operator and community requirements.

2.2 Solution Techniques for Rail Transit System Planning Problem

Several classical and modern optimization methods can be applied to seek the best solution for rail transit system planning with respect to the set of objective functions and constraints discussed in section 2.1. The classical optimization methods include calculus of variation, network optimization, dynamic programming, enumeration, linear programming and numerical search. The non-classical, or modern, optimization techniques include heuristics and evolutionary algorithms, such as tabu search, genetic algorithms and ant colony optimization. During the last few decades, these modern methods have been widely employed to solve station location and rail line alignment optimization problems. This can be attributed to three main reasons.

1- Rail transit system planning is a highly complex multi-constrained problem and the evaluation of candidate solutions can prove both time consuming and challenging, with drastic increases of the number of candidate solutions as the system size increases (i.e., number of lines and stations). Furthermore, the complexity of the problem increases significantly when realistic geographic (e.g., irregular service regions and land use patterns) and demographic (e.g., heterogeneous demand distributions) conditions are incorporated into the evaluation process. Thus, these problems are computationally intractable for realistic rail transit system planning and accordingly they make the use of the modern optimization methods more favourable than the classical methods.

2- The limited ability of the classical methods to handle problems involving significant degrees of complexity (Jha and Oluokun, 2004, Jha and Schonfeld, 2004, Kim et al., 2005, Jha and Kim, 2006, Jha et al., 2006). Attempts to solve highway planning problems using classical methods have revealed these deficiencies explicitly, as shown in table 2.2. Highway planning problems are similar to those in rail transit planning since both share the need to search for a sequential series of spatial elements, satisfy specified geometry requirements and account for topological constraints, as well as land use and environmental impacts. Furthermore, both involve non-differentiable and discontinuous variables, deal with huge amounts of data and require complex computational efforts (Lai, 2012).

3- Specifically, in the context of these data characteristics and demands these modern optimization methods have potential to :

i. Consider and formulate all objective functions and constraints associated with the rail transit system design,
ii. Optimize station locations and rail transit line alignments concurrently,
iii. Automatically avoid restricted areas, and
iv. Be compatible with a GIS database.

Table 2.2: Classical Optimization Methods Used for Highway Planning and their Limitations (Kim et al., 2005, Jha et al., 2006) .

Methods	Limitations
Calculus of Variation	• Requires differentiable objective functions • Not suitable for discontinuous factors • Tendency to get trapped in local optima
Network Optimization	• Cannot yield a smooth alignment • Uses discrete solutions rather than a continuous search space • Requires a large memory
Enumeration	• Not suitable for non-linear cost functions • Inefficient
Linear Programming	• Not suitable for non-linear cost functions • Only covers a limited number of points for gradient and curvature constraints
Dynamic Programming	• Cannot yield smooth alignments • Not suitable for continuous search spaces • Requires independencies among sub problems
Numerical Search	• Tendency to get trapped in local optima • Complex modelling

Among the modern optimization techniques the genetic algorithms (GAs) are the most popular and have a diverse range of applications, particularly in transportation optimization problems such as highway alignment, rail line alignment, rail transit station location and transit route network design and scheduling. GAs are adaptive evolutionary search techniques based on the principles of natural selection and survival of the fittest (Coley, 1999). They treat the search space of a problem as an environment and a set of potential solutions to the problem as a population of chromosomes. This population of chromosomes is evolved over a series of generations to converge towards an optimal solution to the problem. In each generation, the fitness of every chromosome is evaluated with respect to the objective function of the problem, and, stochastically, a number of chromosomes are selected from the current population based on their fitness. The selection process embodies the principle of "survival of the fittest"; that is chromosomes with low fitness values tend to die off, whereas chromosomes with high fitness values tend to survive. The selected chromosomes are then modified (recombined and possibly mutated) with genetic operators- crossover and mutation, to produce a new population. Thereafter, the newly generated population is used in the next generation of the algorithm. After successive generations, depending on the efficiency of the genetic operators (i.e. how well-designed they are), the population will converge towards an optimal solution to the problem (Jong and Schonfeld, 2003) .

GAs have a number of advantages which make them more preferable and widespread search algorithms compared to the other modern optimization methods (Tabassum and Mathew, 2014). One of these advantages is that GAs are less complex than other algorithms. Also, they are easier to transfer and apply in to different real world situations and are therefore more flexible compared to other algorithms. Furthermore, GAs solve both discrete and continuous search space problems very efficiently and with reasonable confidence. For example, ant algorithms have the inherent limitations of being effective only in discrete search spaces, and thus when applied to continuous search spaces they require the search space to be sufficiently discretized (Samanta and Jha, 2012). Moreover, GAs have the ability to perform exhaustive searches and converge towards global optimum solutions. Although there is no rigorous proof to explain why GAs converge toward global optima, several hypotheses have been developed to provide a theoretical explanation for the effectiveness of GAs. It has been demonstrated that GAs work very successfully in many practical applications (Samanta and Jha, 2012, Jha et al. 2006, Jong and Schonfeld 2003, Jong, 1998).

It should be mentioned that, like other classical and modern optimization methods, GAs have limitations. GAs require a large number of responses, (i.e. fitness function evaluations), depending on the number of individuals (chromosomes) and generations, and therefore they can take a relatively long time to evaluate individuals when solving complex problems with a large numbers of parameters. In addition, the success of GAs largely depends on:

(i) The initial population used. An inappropriate initial population can increase the time that the algorithm takes to reach the best solution and, in the worst case scenario, it can prevent the acquisition of a best solution. This can be managed, however, by applying special conditions to the initial population instead of creating it randomly.

(ii) Both the structures and the parameters of genetic operators, in particular crossover and mutation operators. The genetic operators control the search process and GAs' performance in efficiently converging towards the best solutions. In general, these operators are problem specific and, therefore, there is no fixed optimal structure and parameter (i.e., crossover and mutation probability rates) that can be generalized for most real world problems. Accordingly, extensive analysis is required to determine the structure and parameter of these operators that best fit the problem framework under investigation.

2.3 Models for Optimizing Rail Transit Station Locations and Associated Line Network

Since the late 1960s many researchers and planners have made attempts to develop optimization models and application tools for rail transit system planning. These models provide a structured representation of the mathematical formulations of the objective functions and constraints discussed in section 2.1 and the solution techniques described in section 2.2 at a fairly aggregated level that can be used as a tool for a decision-making strategy to come up with a unique and objective solution to the problem. The existing models of rail transit system planning can be broken down into three categories. The models in the first category optimize locations of stations along predetermined rail line alignments. This is since these models are developed based on the assumption that the rail line alignments are determined and fixed prior to the determination of the associated station locations. On the contrary, models in the second category optimize rail line alignments through a predetermined set of stations. That is, these models are developed based on the assumption that the locations of stations are designated and fixed prior to the rail line alignment optimizations. The third category of models, which are sometimes referred to as integrated optimization models, seek optimal locations and sequences of stations without knowing the line alignment, and subsequently apply either linear connections between the stations or conduct a separate optimization process to determine the line alignment in between the stations. The rest of this section reviews the existing rail transit planning models in these three categories.

2.3.1 Category I- Models for Optimizing Rail Transit Station Locations

Early studies in this category concentrated on the interstation spacing along a predetermined rail transit line. Vuchic and Newell (1968) developed an analytical model to determine spacing between stations along a given rail line. The model considers multiple parameters, such as population distribution along the line, access speed to stations, standing time of the train in stations, intermodal transfer time at stations and the dynamic characteristics of the train. However, this is done within a context of a specific case in which the population of an area served by the line commutes to one central point. With the objective of minimizing the passenger travel time, the model calculates interstation spacing by solving a set of simultaneous difference equations specifying the optimality condition. The spacing is considered to be a function of the ratio between the number of passengers travelling on the train and those wanting to board or alight. The model does not consider competitive transport modes.

In a subsequent study, assuming a uniform distribution of passengers along the railway line, Vuchic (1969) proposed a similar model but with a different objective to account for competitive transport modes. This is done by considering a continuous transportation system (e.g., highway) running parallel to a railway line classified as a discrete transport system. It is assumed that the highway has a constant speed and can be accessed at any point along the line as opposed to the railway line, which has discrete movement and boarded only at stations. The model calculated the optimal spacing between stations with the objective of a maximum number of passengers using the railway line assuming that the passengers selected the system on the basis of shorter travel time. The model used a solution method similar to the previous basic model (Vuchic and Newell (1968) and also concentrated on the same case where the population of an area served by the railway line commutes to one central point. It should be noted that these assumptions do not accurately reflect the practical aspects of rail transit planning.

Subsequent work by Kikuchi and Vuchic (1982) developed a theoretical model to determine optimal station spacing and vehicle stopping policy for a rail transit under different operating conditions. The objectives were minimum travel time and minimum total cost (user cost and vehicle operating cost) while considering factors like passenger volume, vehicle capacity, headway and access speed.

The 1990s saw very little progress on the problem of rail transit planning. In 2002, (Laporte et al., 2002), developed a model for locating a prefixed number of stations on a predefined rail transit alignment with the objective to maximise population coverage, subject to a station spacing constraint. The model estimates the demand for each potential station by triangulation of census tracts, assuming that the percentage of the captured travellers decreases with their access distance. To calculate the demand of the neighbouring potential stations, it is assumed that travellers always select the nearest station. The model determines the station locations along the alignment that maximize the objective function using a longest path algorithm on a cyclic graph, containing only the links between the pair of candidate stations that meet the station spacing constraint. In another study, Schöbel et al. (2002), presented a model to locate additional stations along existing railway line networks. The study considered the trade-off between positive and negative effects of additional stations. The negative effect of longer passenger travel times caused by the additional stops made by the train was compared to the positive effect of shorter access time given by additional stations. The objective of the model is to minimize the number (cost) of additional stations while ensuring coverage of all demand centres, assuming the traffic load for each railway line was given. The demand centre, which may represent a settlement area, shopping centre, school, etc., is assumed to be covered if the next station is within a specified radius from it.

In an extended effort, Schöbel (2005) extended his model to a bi-objective model where the maximization of demand coverage is considered in parallel with the minimization of the number of stations. Hamacher et al. (2009) also extended the model by Schöbel et al. (2002). They proposed two objective functions. In the first objective, the number of stations was minimized such that all demand centres would be within a specified radius. In the second objective, the sum of the distances between demand centres and their closest stations is minimized while the number of new stations is fixed. Following the same approach, Carrizosa et al. (2016) proposed a model seeking a set of additional stations covering all demand centres but, instead of minimizing the number of additional stations, the additional passenger travel time due to the additional stations are minimized. When computing the additional traveling time, the model considers the acceleration and deceleration of the trains in addition to their waiting time at stations.

Some studies seeking the optimal locations of rail transit stations have integrated geographical information systems (GIS) and artificial intelligence-based optimization techniques, such as genetic algorithms and ant-colony systems. For example, Jha and Oluokun (2004), developed a preliminary model based on a genetic algorithm (GA) and geographic information system (GIS) to determine station locations along a given rail line alignment. The objective of the model is to minimize total system cost assuming that all travellers commuted to one central point. The total system cost is a function of passenger, operator and capital costs. The operator costs are assumed to be the cost of operating the trains only, while the capital costs are assumed to be a function of stations' land acquisition, construction and parking facility costs.

In a recent study along the same lines of research, Repolho et al. (2013) presented a mixed-integer optimization model to determine the optimal location and number of stations along a planned railway line to be introduced over an existing transport network. The model selects stations within a set of possible locations defined *a priori* to maximize the possible travel cost savings made by the introduction of the new railway line. The model considers the sensitivity of rail usage to time losses due to stops at intermediate stations, as well as competition with the existing modes using the existing transport network. Travel demand is estimated using origin-destination matrices taking into account travel cost, which is more appropriate and realistic than using the population covered by stations, as in (Laporte et al., 2000, Laporte et al., 2005, Gutiérrez-Jarpa et al., 2013)) (see section 2.1.1.4).

Although the review of the aforementioned optimization models has demonstrated the ability of these models to address many aspects of rail transit system planning, they have a number of limitations which constrain their application to real world planning practice. In general, these limitations include:

1- A tendency to overlook the complex interaction between station locations and line alignments by assuming that the rail line alignments are determined and fixed prior to the determination of the station locations. Rail transit stations and the associated line network represent the principle infrastructure of the rail transit system responsible for collecting and distributing passengers and their transportation over distance. Therefore, they underpin the concept of rail transit system operation, and it is vital that they are integrated into a single planning platform in order to achieve optimal (reliable and cost effective) rail systems. Furthermore, assuming rail line alignments are fixed prior to the determination of station locations limits the application of these models only to a very specific set of circumstances in real world planning practice.

2- A focus on a single rail line system and a lack of consideration of systems with multiple lines. This oversimplification of the problem by assuming that the rail line alignment of the system is determined prior to the determination of station locations and that the system consists of only a single rail line further limits the practical applicability of these models.

3- Many of these models assume that the candidate sites for rail transit stations can be located anywhere in the study area or along the rail transit line (Vuchic and Newell, 1968, Vuchic, 1969, Laporte et al., 2002) without considering the various geometric, environmental, topological and budget constraints, which may result in creating many practical issues in real world planning practice. These practical issues include locating stations in environmental restricted/infeasible areas or in areas that may result in significant unnecessary increases in the system cost.

2.3.2 Category II- Models for Optimizing Rail Line Alignment between Pre-set Stations

Compared to the previous rail transit system optimization model category, there are relatively few studies in this category. This can be explained by the fact that rail stations represent the only point at which passengers can have access to a rail transit system, and that it is an important factor for a rail transit system to be selected as an alternative transport mode. Therefore, most rail transit system planning studies focus on the rail stations rather than the line alignment.

The most prominent study dealing with rail line alignment from an optimization perspective was that of Jha and schonfeld (2007). They adapted highway alignment optimization models developed by (Jong, 1998, Jha, 2000, Kim, 2001) to optimize a rail transit line alignment between a pair of stations. The solution method is largely the same but the objective functions and constraints are adjusted to reflect the rail line alignment design criteria. The objective functions of the model are to minimize passenger and operator costs while satisfying the general geometry constraints of rail line alignment (such as horizontal and vertical curve radius). The passenger cost is formulated to involve the cost of access time, riding time and waiting time. The operator cost is formulated to involve track-related construction costs, right-of-way costs, earthwork costs and rail operating costs. In the model, GIS is integrated with GA to perform the optimal search. The GIS provides geographical data such as land values, existing road networks and topography to GA which ultimately generates and evaluates the candidate solutions.

Similarly, using a heuristic solution algorithm based on GA integrated with a background GIS database, Lai and Schonfeld (2010) proposed a model for optimizing a rail line alignment that can connect several predetermined stations. With respect to the objective of minimum construction cost, the proposed model generates the alignments through pre-set stations while satisfying the geometry constraints of the rail line alignment. The geometry constraints incorporated in the model included the general geometric requirements for the horizontal alignment and the vertical alignment. In an extended effort, Lai and Schonfeld (2012) extended this model to incorporate additional objectives into the optimization framework. The additional objectives are passenger and operation costs. Furthermore, vehicle dynamics, which account for both the horizontal /vertical alignment and the station spacing, are considered when formulating both the operational and passenger costs. With the objective of a minimum total cost, which is the sum of the construction, passenger and operator costs, the model generates the alignment between the predetermined set of stations while satisfying the geometry constraints along the alignment.

Along the same lines, integrating a heuristic solution algorithm based on GA with a GIS background database, Kang et al. (2014) proposed a model for optimizing rail line alignment between a predetermined set of stations. The objective of the proposed model is to minimize total cost, which consists of the three cost items of construction cost, life-cycle cost and penalty cost. . The life-cycle cost comprises the cost items that occur throughout the life cycle of the system, such as structure maintenance cost (e.g., bridges and tunnels), passenger cost (in train travel time, waiting time and access time) and train operation cost. The penalty cost is used to control the geometry requirements along the generated line alignment. That is, the model applies the penalty cost if the generated alignment violates the geometry constraints, which include minimum horizontal and vertical curve radius and minimum and maximum gradient constraints.

In a recent study, Costa et al. (2013) presented a heuristic optimization model based on a simulated annealing algorithm to optimize a high speed rail (HSR) alignment that minimizes an objective function considering construction costs while complying with stringent geometry, land use and location constraints. The objective function consists of five terms: construction costs; a penalty value for gradient noncompliance; a penalty value for horizontal angle noncompliance; and a penalty value for land-use noncompliance so as to restrict the HSR configuration across restricted areas. A location benefit is included so that the HSR line connects all the mandatory locations specified by the planners. In the model, it is considered that the penalty and benefit coefficients are to be established through expert judgment taking into account the problem specific. It is also assumed that whereas in reality the three dimensional (3D) configuration of the rail alignment is defined by a set of tangents and curves, both in the horizontal and in vertical planes, this model uses linear sections that connect a set of sequential 3D points in space. Costa et al. (2016) improved this model by instituting a solution technique that was calibrated for a simple and synthetic case study in order to solve a large scale problem size and by presenting the application of the model to a real world case study in Portugal based on real data.

In summary, the review of the above-mentioned models in this category sets out their capability to consider various local factors, such as topography, land use pattern and land value and various geometry constraints for evaluating alternative solutions. However, the models have many limitations that should be addressed for them to be viable in real world applications. These limitations are:

1- Similar to the models presented in the first category, the models in this category neglect the complex interaction between station locations and line alignments, and focus on a single rail line system. The station locations are determined or prefixed prior to the determination of rail line alignments.
2- The model by (Jha et al., 2007) is capable of generating rail line alignments that can only connect a pair of stations and as a result, can be applied only to situations where there are no intermediate stations between the line terminals; these scenarios are seldom found in real world planning practice.

2.3.3 Category III- Integrated Models for Optimizing Station Locations and Associated Lines

As discussed in section 2.3, this category attempts to arrive at optimal locations and sequences of rail transit stations and determines line alignments in between stations either through the linear connection of stations or in separate optimization process. Early studies in this category include the work of Dufourd et al. (1996), which addressed the problem of locating a rapid transit line with a known terminus linking a fixed number of stations on a grid network. With respect to the objective of maximum population coverage, the model determines station locations while the line alignment is generated according to the shortest path between the stations. The population covered by the station is calculated using the station catchment area approach described in section 2.1.1.4, and on the assumption that stations do not have overlapping sections. A Tabu search heuristic was developed to explore the solution space for the best value of the objective function while ensuring the station spacing constraints were met. Similarly, with respect to the objective of maximum population coverage and interstation spacing, Bruno et al. (2002) proposed a two-phase heuristic model for the location of a single rapid transit alignment. The first phase generates an initial alignment through gradual extension from a station to the consecutive station with the aim of yielding maximum population coverage while respecting the station spacing constraint. The second phase improved the initial alignment by extracting from it a partial alignment of a specified number of consecutive stations and extending this into several full alignments using the same procedure as in the first phase. The population coverage is calculated using the station catchment area approach described in section 2.1.1.4, which was also used by Dufourd et al. (1996). Since the population coverage does not reflect the primary objective of rail transit system planning, however, which is improving population mobility, as discussed in section 2.1.1.4, a number of studies diverted from attempting to capture the maximum population coverage to traffic coverage. This can be seen in many studies (Laporte et al., 2005, Lai and Schonfeld, 2012, Lai and Schonfeld, 2016).

Laporte et al. (2005) study proposed a greedy algorithm for locating a single rapid transit line alignment through a set of stations in order to maximize total traffic flow subject to a length constraint. The proposed model first derives a trip coverage matrix for each station pair by combining the notion of measuring the population coverage of a potential station as proposed by Laporte et al. (2002) with census data. Each element in the matrix represents the demand caught by the stations between each origin-destination (OD) pair. The model then applies a simple logit model to calculate OD demand via the rail transit. The approach used for estimating station-to–station origin-destination demand, however, is neither validated nor had a theoretical base. Lai and Schonfeld (2016), proposed a heuristic based on GA and GIS to determine rail transit alignment and associated station locations connecting a pair of a predefined terminal stations simultaneously. The model is formulated to take into account both the maximization of trip coverage and the minimization of the system construction cost while satisfying the general geometry requirement of the line alignment and station spacing constraints. The objective is to minimize of total system cost, which is a function of the system construction cost, passenger and operational costs. Both the passenger and operational costs are formulated to maximize the total trip coverage, for which a multinomial logit model is used to estimate station-to-station demands.

In another study, Samanta and Jha (2008), presented a two-stage analytical model to optimize the station locations for a single rail transit line. The first stage, embedded within GIS, identifies feasible station locations to avoid interference with existing road networks and built-up areas, such as major residential, commercial and business localities. The second stage applies an optimization procedure based on GA to obtain optimal locations and sequences of stations from the feasible location set identified in the first stage, assuming that the line alignment constituted a linear connection of the stations. The objective of the optimization is to minimize total system cost per person, which is a function of passenger travel time cost, system operation cost and construction cost (here, only the stations' right-of-way costs), while ensuring that station spacing remained constrained. However, the model does not consider demand estimation even though this is a very important component for rail transit system planning. Accordingly, Samanta and Jha (2011) built upon their original study by seeking to satisfy further objective functions of demand and cost. In the model, the methodology remains mostly the same, but the objective functions are extended to three different objective functions of demand and costs. The first objective is to minimize total system cost per person, which is a function of passenger travel time cost, system operation cost, and construction cost (again, only the right-of-way costs). The second objective is to maximize the demand covered by the rail stations. The passenger travel time cost per person is minimized separately as the third objective function. With the number of stations and spacing between stations constraints, the model optimizes these three objectives individually for the best solution. The decision-making process for choosing the best solutions out of the three different objective functions is assumed to be a judgment based on the existing situation, nature of the transportation problem and location of study area.

Along the same line, a number of studies focused on the problem of determining multiple rail transit lines and associated station locations using similar criteria to planning a single rail transit line system. Bruno and Laporte (2002) improved the heuristic model introduced by Bruno et al. (2002) to locate multiple rapid transit lines possessing a given topological configuration. The shape of the line network to be built, a corridor within which each line is to be located, the number of stations along each line and the station spacing constraint are assumed to be specified first by the model user. The model then applies a Tabu search heuristic to construct each alignment within each of the defined corridor. Using integer programing, Schöbel and Scholl (2005) presented several models to select a subset of lines for a line pool to connect several stations so as to minimize passenger inconvenience under a budget constraint. The passenger inconvenience is measured by the travel time and the number of transfers along the lines, and the cost of each line is assumed to be known beforehand.

Laporte et al. (2007) presented a mathematical model to determine station locations and to generate a network of lines linking the stations. Assuming a fixed origin and destination of the line network, the model is formulated to maximize travel demand as an objective function and the construction cost as a constraint. This model was extended by (Marín, 2007) to incorporate the line location constraints with a bounded but variable number of lines, and lines with no fixed origins and destinations. However, in both models travellers are assigned to the rapid transit mode if the associated cost of using the network is less than or equal to the corresponding cost of the competing modes.

In a more recent study, Laporte et al. (2011) formulated rail transit network planning as a dual criteria problem in which the two objectives are the construction cost and the population covered by the network. The model is designed to minimize a linear combination of these two objectives that is achieved by subtracting from the construction cost the population covered by the network multiplied by a positive parameter, which is controlled by the model user. This parameter is introduced in the model to control the fraction of the population covered by the network. The higher the parameter value the larger number of population are covered by the network. Similarly, Gutiérrez-Jarpa et al. (2013) modelled and solved the problem of locating a rapid transit network in such a way as to maximize origin-destination traffic capture and minimize construction cost. A mixed integer formulation is presented to solve the problem, assuming a predefined topology of the network. The model locates stations within corridors associated with the various segments of the proposed topological configuration as well as generating rail transit lines to link the stations while satisfying station spacing constraints. To deal with both objectives, the problem is first solved for each objective separately then a linear combination of the two objectives is minimized, corresponding to the straight line passing through the extreme solutions just found. The construction cost of the network lines is calculated as a linear function of their corresponding length. The captured traffic demands were computed in the absence of other competing transport modes, where a traffic demand was considered to be captured only if both its origin and its destination lay within a pre-set distance of their respective stations.

Although models in this category have attempted to address some of the limitations identified in the first and second model categories either through consideration of complex interactions between rail transit station locations or consideration of rail networks with multiple lines, many limitations remain. Therefore, further improvements are needed to arrive at a reliable and cost effective solution. These limitations are:

1- The definition of the rail line alignments by a linear or the shortest path through the optimized stations without consideration of many crucial factors, such as the cost of land acquisition and construction, as well as land use patterns and environmental impacts (Dufourd et al., 1996, Bruno et al., 2002, Laporte et al., 2005, Samanta and Jha, 2011). Therefore, the models do not adequately capture the complex interaction between the two interdependent elements of rail transit systems (i.e., line alignments and associated stations).
2- A singular focus only on either the coverage of traffic demand or population (Dufourd et al., 1996, Bruno et al., 2002, Laporte et al., 2005), thus disregarding the other crucial aspects of rail transit system planning such as construction costs. In real world practice the attempt to capture both maximum population / traffic trips and construction cost is crucial. In general,

higher investment in building rail lines would deliver either higher service quality for passengers or lower operational costs for operators, or a combination of the two. Yet, these benefits may or may not be worth the additional investment cost. Therefore, it is vital to apply a trade-off between the capital cost and both service quality and operational productivity.

3- Some of these models recognize the importance of the integration of the construction cost into the optimization framework (Schöbel and Scholl, 2005, Laporte et al., 2011, Gutiérrez-Jarpa et al., 2013). However, they neglect the effect of the various local conditions of the study area, such as land use patterns, land values, soil conditions and topography on the construction cost of the system. Instead, they simply assume that the construction cost of the system is a linear function of the line network length and number of stations. However, in reality the construction cost of rail transit not only depends on the line network length but also on the locations of both stations and line alignments. For example, tunnel cost which comprise a large share of the system's total construction cost is largely influenced by the location of the line alignment, i.e., whether the line alignment passes through soil or bedrock. The construction cost of a rail transit system cannot be captured by a linear function and this assumption therefore, calls into question the ability of these models to arrive at optimal solutions.

2.4 Summary and Research Contribution

This chapter has presented the formulation of rail transit system planning within the context of an optimization problem, elaborating various objective functions and constraints that are associated with rail transit system planning. This chapter has also presented a review of the various solution techniques that can be employed to solve the problem. A special focus has been given to the characteristics of these techniques in order to provide insights into their efficiency in determining optimal solutions. With respect to optimization methods, many specific models have been with a particular focus on the interlinked aspects of rail transit system planning; determination of station locations and line alignments. The review has classified these models into three main categories based on the problem formulation and methodology used. Despite their merits in certain aspects, many limitations have been identified that necessitate further improvements in order to achieve optimal (reliable and cost effective) solution to the transit system planning problem.

Most of the existing models, particularly those which fall in the first two categories(i.e rail transit station locations and rail line alignment between preset stations optimization models) , ignore the complex interactions between rail line alignments and station locations, addressing either line alignment or station location optimization but not both. Models classified under the first category, focus only on the optimization of the station locations assuming that the line alignments are predetermined or prefixed. In contrast, models that fall in the second category focus only on the optimization of the rail line alignment assuming that the locations of the stations are prefixed. Models in the third category optimize the locations and sequence of stations and determine line alignment either through linear connections of the stations or in a separate optimization process. These approaches, although widely used in rail transit planning practice, cannot effectively capture the interactions between rail line alignments and station locations. Therefore, there is a risk that the resulting solutions from these models are not optimal. The network of lines and associated sets of stations represent the interdependent principal infrastructures of a rail transit system and coordinating them into an integrated optimization model is crucial to the acquisition of an optimal (reliable and cost-effective) rail transit system.

A review of multiline and associated stations models has revealed that they ignore crucial decision variables (such as land use patterns, land values , geology and topography) in order to simplify the computational burden, compromising on reliability and cost effectiveness.

In order to address the limitations of the existing studies identified throughout section 2.3 and summarized in this section, this research develops an integrated optimization model that can effectively:

1- Determine rail transit station locations and line network alignments connecting the stations while satisfying the various real world constraints of rail transit system planning.
2- Consider the complex correlations and interactions between the rail transit stations and associated line alignments by integrating these two intertwined elements into a single optimization process.

3- Account for various local conditions and requirements for passenger convenience, operator productivity and community benefits, and perform trade-offs between them. This trade-off is essential to the acquisition of a rail transit system that can provide high service quality and benefits for both the passenger and the whole community, with an acceptable capital and operational costs.

The next chapter develops a methodology to address the limitations of the state-of-the art identified in this Chapter.

Chapter 3 - A New Framework for Rail Transit System Planning

Based on the limitations identified in the existing models and the objectives of this research, this chapter develops a new methodology for rail transit system planning. The proposed methodology is designed to incorporate the complex correlations and interactions between the rail line alignments and station locations by integrating them into a single planning platform. Furthermore, it takes into account various local factors, including travel demand patterns, socio-economic growth, topography, existing road networks and soil conditions, as well as the constraints of transit rail system planning.

This chapter details the assumptions and formulations of the methodology across five main sections. Section 3.1 presents an overview of the methodological framework to obtain a rail transit system that can best fit the desired objectives. Section 3.2 identifies the various requirements that must be considered in rail transit system planning and discusses their relative values and significance. Section 3.3 identifies the factors upon which the requirements identified in section 3.2 depend and translates them into effective criteria that can be used in the evaluation process. Section 3.4 proposes a comprehensive procedure to simultaneously determine station locations and the network of lines connecting them based on the requirement criteria identified in section 3.3 within the context of an integrated optimization process. A summary of the new methodology is presented in section 3.5.

3.1 System Development Framework

Station locations and the network of lines connecting them are the principal elements of a rail transit system. The planning of such a system can be conceptualized as either a single line or a network of lines connecting the stations. It can also be planned as either an entirely new system or an enlargement and improvement of an existing rail transit system, depending on the traffic congestion growth rate, and the level of demand for public transport in the area to be served. The proposed methodology encapsulates a comprehensive procedure that can be used efficiently for the planning of a new rail transit system and the expansion of an existing one, based on trade-offs between various rail transit system requirements and constraints.

The designed methodology consists of three main levels of analysis and decision-making processes, as illustrated in figure 3.1. Level I identifies the requirements that must be accounted for in rail transit system planning. This involves the consideration of the level of passenger service, operator productivity and potential benefits for the community. The analysis and decision-making process at level II includes the translation of the requirements of the three stakeholders identified at level I into quantified system requirements that describe the characteristics and performance of the rail transit network efficiently. Level III establishes a comprehensive procedure for generating a family of good solutions and then evaluates these within the context of an integrated optimization process. The analysis at level III resolves the essential trade-offs between an effective rail system that provides high service quality and benefits for both the passenger and the whole community, and an economically efficient system with acceptable capital and operational costs. It also determines the role and impact of each stakeholder preference on the final configuration of the system. The details of each of the decision-making levels are discussed in the next three sections.

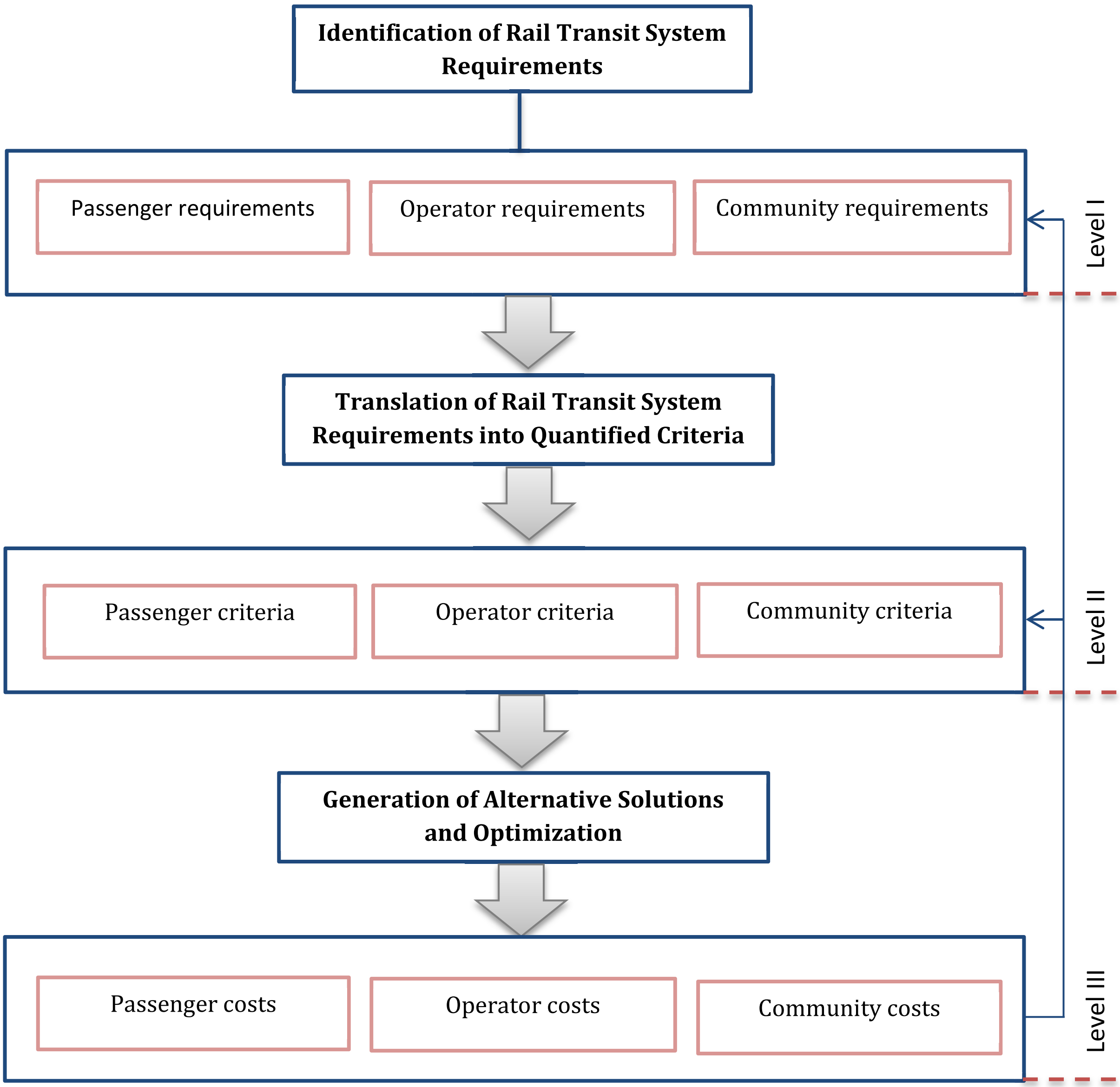


Figure 3.1: Proposed Methodological Framework

3.2 Level I – Identification of Rail Transit System Requirements

The network of lines and the stations associated with these lines represent the principal infrastructures of a rail transit system and underpin the concept of operation that involves the collection/distribution of passengers and their transportation over distance (Vuchic, 2005). The decision regarding where to place rail stations and how to connect them in a network of lines not only determines the operational and network characteristics of the planned system, but also influences the role the planned system plays in the development of the served area's physical form, economic activities, and social and environmental conditions. These two interlinked elements (i.e. station locations and network of lines) of the rail transit system, therefore, represent crucial parts of the planning process.

If the desired objectives of the rail transit system are to be achieved, a range of requirements must be considered in the planning process. These requirements, which may have different relative importance in different cities or served areas, can be classified into three main groups; passengers, operators and community. Each of these groups has a set of complex interrelated requirements which may overlap or contradict with those of other groups. If these relationships are not clearly defined and distinguished from each other the decision-making process is likely to become much more complex. Consequently, the following subsection details the main requirements of each of these three interrelated groups, as well as their significance.

3.2.1 Passenger Requirements

Passengers utilize a rail transit system as a transport mode with the system mainly built for this purpose. Their main objectives in using a rail transit system are to arrive at a desired destination quickly, comfortably, safely and at a reasonable cost. These objectives accommodate a broad range of requirements including availability of stations, travel speed along the rail line network, comfort, punctuality, safety at stations and on board, integration with other transport modes and the cost of utilizing the system services (Vuchic, 2005). These requirements are very diverse, with some significantly influenced by locations of stations and line network configuration, while others are largely influenced by other aspects of a rail transit system, such as the design and manufacturing quality of the system and its components and the service quality and operational reliability of the system. For example, the major elements influencing passenger comfort during an entire rail trip are availability of seats, the dimension and design of seats, the load factor, train vibration, width of train carriage doorways, a noise, temperature, humidity and ventilation inside the train (Kardas-Cinal, 2010, Howarth et al., 2011).

Punctuality can be defined as the provision of rail services at scheduled train arrival and departure times . It is measured as the percentage of trains that arrive at or depart from the desired destination within a certain interval after the scheduled time. Punctuality largely depends on the reliability of the rail system infrastructure, signalling system and rolling stock, as well as the number of passengers and the occupancy ratio (passengers/seats) (Olssona and Haugland, 2004, Goverde, 2005). Safety refers to the absence of incidents that may lead to injury to persons and/ or physical damage (Vuchic, 2005). Safety at rail stations is largely influenced by station design and layout, including geometry and visibility of steps, signage and flooring materials, as well as easy boarding of trains and control of train dispatch from stations. The main elements of safety on trains include the characteristics of the trains, the design of the routes over which the trains move, and availability of signalling systems that ensure a fully protected path for the trains through switch and crossing points (RSSB, 2011, ORR, 2007). The characteristics of train include the presence and functionality of relevant safety systems and equipment, including braking systems, internal and external lighting systems, door safety systems, emergency equipment and signage, fire detection systems and equipment monitoring systems. Station availability, travel speed and integration with other transport mode requirements are strongly linked to the locations of rail stations and associated line network configurations. The following subsections describe these requirements and detail how they are influenced by the locations of rail stations and the associated line network.

3.2.1.1 Station availability:

This can be defined as the distance to stations that should be reasonably close both from passengers' origin and destination points. This element is very important both for passengers' trip travel time and their comfort. For good station availability, passengers require stations with close proximity. A number of empirical studies have been carried out to identify the significance of the distance to and from stations for passengers' convenience and how this affects their decisions as to whether to use rail (Reilly and Landis, 2002, Kitamura et al., 1997, Samanta et al., 2005, Brons et al., 2009, Cervero, 2001). The findings of these studies highlight the importance of the proximity of rail stations to passengers' satisfaction with the rail service provided and therefore confirm that railway passengers view proximity as an important determinant in their choice of rail as an alternative transport mode. However, the degree of importance attached to station proximity depends largely on the transport modes that passengers use to access rail stations. It holds more significance when walking is the dominant access mode compared with private vehicles or bus modes (Cervero, 2001). This is because a long walking distance to and from stations usually tends not only to increase travel time but also affect passenger comfort, especially in an unpleasant environment or bad weather (Vuchic, 2005, Loutzenheiser, 1997, Cervero, 2001).

3.2.1.2 Travel speed

Generally, travel speed can be defined as a unit distance that trains can travel along the rail network within a unit time while meeting specified safety, comfort, economy and train performance requirements (Vuchic, 2005). Travel speed on a rail network is the most important characteristic influencing both the passengers' travel time and their mode choice decisions. Increasing speed results in shorter passenger travel times and increased passenger satisfaction levels, and thereby increased rail patronage. This implies that travel speed, in addition to the level and quality of service experienced by passengers, is strongly linked to rail system performance and operation, which will be discussed in more detail in the operator requirements section. Travel speed is therefore viewed as a major decision variable in the planning of rail transit systems to guarantee a good service quality for passengers (Jha and schonfeld, 2007, Samanta and Jha, 2011, Lai, 2012).

3.2.1.3 Integration with other transport modes

This refers to the degree to which a rail transit system is connected and coordinated with an existing transport network, such as the road network, airport terminals and ports. A railway journey, from the passengers' perspective, is part of a journey chain that includes a journey to, and later from, the rail station using different modes of transport (Givoni and Rietveld, 2007). The passengers' satisfaction level with the rail journey can therefore, be considered not only to be a function of the level and quality of the rail service but also the level and quality of service on the transport modes to and from the rail stations. Research over the last few decades has explored how the integration and coordination among different transport modes affects travellers' behaviour and shapes their mode choice decisions. In particular, the research focus has been on evaluating how important the quality of access/egress facilities to the rail stations is to passengers, and the degree to which this affects their overall satisfaction with the rail journey. For example, a case study on the rail network in the Netherlands (Brons et al., 2009) found that passengers' satisfaction with their rail journey was partly the result of their satisfaction with the level and quality of the access service and facilities provided to them. Furthermore, the results showed that the quality of the accessibility, which was represented by the frequency and travel time of the public transport service to the rail stations, and the quality of bicycle and car parking facilities around the station areas, was more important for infrequent rail passengers. This indicates that the quality of access services to the rail station, in addition to passengers' satisfaction with the rail journey itself, affects propensity of traveling by rail. In another study on the rail network in the Netherlands Givoni and Rietveld (2007) obtained similar results and found that improving connections with public transport services and car parking facilities contributed to a higher satisfaction level with the rail journey for those who accessed the station by these modes. Even more significantly, the results showed that some travellers avoided utilizing rail because of their perception of the relatively low quality of the station and its level of accessibility. These findings confirm the importance of ensuring a seamless integration between rail services and other transport network services. Such integration would appear to have great potential for increasing the propensity to rail use and is therefore, a very important aspect in the planning process.

3.2.2 Operator Requirements

"Operator" refers to the entity responsible for managing, operating and maintaining a rail transit system. In general, operators are purely public, private, or local government transport authorities, or sometimes public-private-local government partnerships. Regardless of the sector of the operator, satisfying passengers' basic needs with minimal operational costs is the main objective of the operator. The requirements of this group are therefore interrelated with, but by no means identical to, the passenger requirements. For example, the operator must ensure that each station and line in the network provides adequate service facilities, capacity, mechanical reliability and labour utilization, while passengers are not concerned about these aspects as they prefer services that are fast, reliable, and comfortable (Vuchic, 2005). Station availability, population coverage, passenger attraction, travel speed, capacity, integration with other transport modes, operational reliability and safety can be considered as the basic elements that most strongly affect the productivity level of operations, i.e., the cost of providing a given service quality for passengers. Some of these requirements are largely influenced by the locations of rail stations and the associated line network, while others are affected by other aspects of rail transit systems. For example, capacity, which is a measure of the capability of a rail transit system to transport a specific number of passengers on rail transit lines in a specific time period and under specific conditions, is largely dependent on infrastructure, traffic and operating parameters. Operational reliability can be defined as the ability of a rail system to offer its required level and quality of service. This largely depends on the operational complexity of the system, the design and manufacturing quality of the system and its components, and the maintenance of the system (Vuchic, 2005). Station availability, population coverage, passenger attraction, travel speed and integration with other transport modes are, however, influenced by the locations of stations and the line network configuration. The following subsections describe these requirements and detail how they are influenced by the locations of rail stations and the associated line network. It should be noted that a number of the requirements that are discussed in section 3.2.1 are also discussed in these subsections but from the operator's point of view.

3.2.2.1 Station Availability

The number of railway passengers can be considered to be not only a factor of potential demand but also the rail service offered and the degree to which this fulfils passenger needs. The level and quality of the services to and from rail stations represents an important dimension of the rail service, which influences the travel time and satisfaction level experienced by passengers (Brons et al., 2009, Samanta et al., 2005). From the operator's perspective, this emphasizes the importance of station availability as an effective planning tool to increase system usage, which in turn leads to reduced operation cost. Research has examined how important the access/egress distance to/from stations is to passengers' satisfaction levels and the role this plays in increasing railway patronage. Using land use, road network and public transit information and a household survey of five neighbourhoods in the San Francisco Bay Area, Kitamura et al. (1997) examined the propensity of people to use rail including variables representing access to transportation facilities. The study showed that the total number of rail trips and rail modal shares are both significantly associated with the accessibility to rail stations (defined here as distance to the nearest rail station). A case study on rail networks in the Netherlands revealed that the propensity of people to use the rail system decrease by 20% if the access distance to railway stations exceeds 500 metres (about 5 minutes walking time). This distance decay effect rises to about 30% at distances between 1.0 and 3.5 km, and can reach up to 50% above this distance (Keijer and Rietveld, 2000). In another study on the rail network in the Netherlands, Brons et al. (2009) further highlighted the importance of access distance to rail stations and its potential in increasing rail use. The results showed that decreasing the average distance from the postcodes centroid to the station by 1% (about 100 metres) can lead to an increase of 350 rail trips per year.

3.2.2.2 Population Coverage

This can be broadly defined as the number of people who live within a usual walking distance of rail stations (i.e., 5-15 minutes access distance from rail stations). Generally, the higher the population size, the larger the potential number of rail system passengers within this area and, accordingly, the higher the operator productivity level. The relationship between population density and rail transit ridership has been examined in several studies. For example, Loo et al. (2010) analysed the factors that are expected to contribute to higher rail transit ridership by using multiple regressions, taking the rail systems in the city of New York and Hong Kong as case studies. The results showed that the demographic characteristics of population density/size are among the most important factors influencing rail transit ridership in these cities. When the population size of New York increased by 100 in their model, average weekday railway patronage would tend to increase by 23. In Hong Kong, these demographic characteristics are positively associated with railway patronage, such that doubling the population density would tend to increase the average weekday railway patronage by about three times. These findings are consistent with other results (Cervero, 1993, Cervero and Kockelman, 1997, Cervero et al., 2004, Evans et al., 2007, Frank and Pivo, 1994, Yoh et al., 2003) that confirm that the population coverage is a very important factor in contributing to high rail transit patronage. Therefore, to increase rail ridership and the associated revenue gain it is essential to incorporate this element into the rail transit planning platform.

3.2.2.3 Passenger Attraction

Passenger attraction is difficult to define precisely but, broadly, it refers to the ability of the planned rail transit system to encourage people to utilize it as an alternative transport mode. The number of people who utilize a rail transit system can be considered to be a function of potential demand and the degree to which it meets the passenger needs defined in section 3.2.1. In addition the number of potential passengers can be significantly boosted by ensuring that the system is well-known and easy to use (Vuchic, 2005). This includes aspects such as coordination with land use pattern, particularly high-density commercial land uses (like central business districts (CBD), recreational centres and office complexes), walking environment, mobility and accessibility to activities surrounding rail stations, marketing and the physical image of the system (Cervero, 1993, Cervero et al., 2004, Vuchic, 2005, Evans et al., 2007, Loo et al., 2010). Attracting a large number of people to utilize the planned system plays a key role in raising operators' productivity levels. This is because the greater the number of passenger trips the planned system carries the more economically it operates (i.e., its unit operation cost decreases), and subsequently the higher the revenue gains. Enhancing passenger attraction is therefore, widely viewed as an effective planning strategy for increasing the efficiency of public transit operations.

3.2.2.4 Travel speed

The train travel speed is an important factor not only influencing the service quality for passengers but also attracting passengers and enhancing the economical efficiency of operation of the system. It also influences the number of trains required to provide a certain level of service and hence, the system capacity and operating costs (Vuchic, 2005, Lai, 2012). Compared to low operation speed along a rail network, high operation speed results in lower operation and maintenance costs. There are three main reasons for this (Garcia, 2010). The first is that increased speed makes it possible to travel more kilometres per unit time and therefore reduces the operational costs, which are usually measured in unit cost per passenger-km. The second reason is that reduced train maintenance costs due the larger radius of rail line curvatures on which the trains move results in less usage of air breaks and a lower number of catenary conductors. The third reason is the fact that the high speed makes it possible to charge passengers a higher price which results in higher operational revenue. In order to achieve higher operational speed, however, a longer distance between stations must be considered in addition to wide horizontal curves and low gradients in the line network. Although a long distance between rail stations is very important for improving the operational speed and achieving better system performance, it has obvious risks in reducing the potential number of rail passengers and thus reducing operator revenues. Therefore, careful attention must be paid to the station spacing in order to balance an acceptable operational speed with station accessibility.

3.2.2.5 Integration with other transport modes

The degree to which a rail system is connected and coordinated with existing transport networks, such as the road network, airport terminals, ports and other public transport services is a very important characteristic that influences the operational cost of the system. For the operator, good integration with other transport modes promotes easy transfers for passengers, and shorter out of train time (walking, waiting and transferring). This, in turn, can make the rail system a more attractive travel option and thus, ultimately, contribute to increasing rail usage. Findings from past research on the connectivity between rail and other transport modes, such as those associated with ease of travel to and from rail stations and convenience of transfers during a journey, all indicate that successful integration can be an important planning factor to boost rail transit usage (Sung and Oh, 2011). Even more significantly, the Brons et al. (2009) study on the rail network in the Netherlands found that improving and expanding connectivity between rail and the other transport mode services can substitute for improving and expanding the services provided on the rail network, and might be more cost effective when the aim is to increase rail use. In a case study on the rail network of Seoul, Korea, Loo et al. (2010) lend further support to these studies. While examining the planning factors that have potential to boost rail transit usage, it was found that a better integration between rail and public transport services, particularly between rail and bus services not only contributed to higher rail usage but also led to increases in bus usage through transfers between rail and bus services. These findings, therefore, underscore the importance of integrated transport to not only achieve increased railway patronage, but also make the utilization and operation of both rail and other transport network options more efficient and economical.

3.2.3 Community Requirements

Community refers to the entire population in the area potentially served by the rail system, including passengers and operators. Community objectives are more diverse than the two preceding groups. They involve not only most of the passenger and operator objectives, but also impacts on the population and the environment, as well as the economic activity of the served area. In a broad sense, the main objectives of the community are a rail transit system that is easily available for all residents, with a service quality that attracts passengers, reduces traffic congestion and improves economic activities with minimal environmental impacts and investment costs. These objectives accommodate a broad set of diversife requirements, such as station availability, population coverage, passenger attraction, economic growth, environmental impacts and construction costs (Vuchic, 2005); some of which overlap with the passenger and operator requirements. The following subsections detail the community requirements, discussing those that coincide with the passenger and operator requirements from the community perspective.

3.2.3.1 Station Availability

Station availability can be viewed as the most important single characteristic of rail transit system and is one of the objectives of all interested stakeholders. In addition to its contribution to increasing passengers' satisfaction level and operators' productivity levels, it plays a significant role in coaxing people out of their cars and into trains, mitigating traffic congestion and enhancing mobility. Basically, a shorter distance between stations and passengers' origin and destination locations results in a reduced out of train travel time, increased passenger convenience and thereby increased system usage. This in turn leads to a decrease in road vehicle trips and thus reduced traffic congestion and emissions. As discussed in sections 3.2.1.1 and 3.2.2.1, researchers have studied the access distance to railway stations and the degree to which this influences passenger satisfaction and operator productivity levels. In addition, some researchers have examined the significance of station availability in the mode choice between car and rail. For example, Stringham (1982) examined variations in rail mode share for all trips generated by a land use, as a function of its distance from a rail station using two Canadian rail systems as case studies; the Toronto subway system and the Edmonton light rail system. The study found that within a radial distance of 3000 feet (about 900 metres) from a station, rail transit mode share ranged between 30-60 percent of all commuting and education trips. The Stringham study also conducted further analysis to examine how access modes of rail transit users vary with distance from a station. The results showed that over 90% of rail usage whose origin or destination was within 1500 feet (about 450 metres) of a station walked to the station.

At a distance of 3200 feet (about 975 metres) bus transit was the dominant mode of access. At 3700 feet (about 1200 metres), the access mode share were; 0% for walking, 15% for car and the reminder arrived by bus transit. In another study, Cervero et al. (2004) found a similar pattern across 129 rail stations in the San Francisco Bay Area using the 2000 US census data and geographical information system (GIS) tools. For residents, the average transit mode share within 0.5 miles of a station was 27% compared to 7% mode share for residents between 0.5 miles and 3 miles of the station. For office workers, average transit mode share at 0.5 miles distance was 19% as compared to 5% region wide. These findings underscore the potential mitigation of traffic congestion and, by extension, the environmental benefits that could occur from providing a good station availability for the proposed rail system.

3.2.3.2 Population Coverage

From the community point of view, maximizing population coverage is a very important aspect in rail transit system planning because it strongly influences the system usage. Increasing population coverage leads to an increase in the number of potential passengers utilizing the system, which in turn results in reduced overall vehicle travel, and thereby reduced traffic congestion and improved mobility. Several research studies have examined the effect of built environments on travel behaviour and the degree to which it stimulates rail system usage and reduces dependence on cars (Cervero, 1993, Cervero and Kockelman, 1997, Evans et al., 2007, Loo et al., 2010, Frank and Pivo, 1994, Hong et al., 2014, Yoh et al., 2003). The results of these studies confirmed that population density is one of the important predictors of mode choice and can significantly boost rail transit usage and reduce car dependence. For example, using the 1996 household travel survey, Reilly and Landis (2002) analysed the effects of population density on individual mode choice decisions made by travellers in the San Francisco Bay Area by constructing a series of multinomial logistic regression models of mode choice for a variety of non-work trip purposes. They found that population density plays an important role in influencing mode choice decisions and reducing reliance on cars. On average, an increase in the average density of four persons per acre within one mile of an individual's residence is associated with a 7% increase in the probability of walking or taking transit. In a subsequent study of the San Francisco Bay Area rail transit system, using the 2000 US census, Cervero et al. (2004) found that a doubling of residential densities from 10 to 20 dwelling units per acre leads to a 3.7% increase in rail transit's commute mode share. Chen et al. (2008) evaluated the impact of density in influencing people's mode choice decisions in the New York Metropolitan Region. Their results confirmed the important role that the population density plays in shaping the people's mode choice decision and propensity to use cars. Overall, these findings indicate that incorporating population coverage into the planning platform can be viewed as an effective strategy for increasing the efficiency of the planned system, decreasing the reliance on cars, and subsequently reducing traffic congestion and emissions.

3.2.3.3 Passenger Attraction

Attracting people to use a rail system is not only important for increasing operator revenues but also essential for increasing the usage of the planned system, reducing dependence on cars, and subsequently reducing traffic congestion and associated emissions. The higher the frequency of a service, the higher the number of passengers carried. This reduces travel time and the need to use cars. This in turn serves to reduce the overall vehicle kilometres travelled and associated traffic issues. Many researchers have studied the factors influencing the attractiveness of public transit systems to people and the degree to which attractiveness contributes to coaxing people out of their cars and into trains (Cervero, 1993, Cervero et al., 2004, Evans et al., 2007, Cervero and Murakami, 2008, Loo et al., 2010, Sung and Oh, 2011). The findings of these studies indicate that integrating rail systems with land use patterns, particularly, high density residential, commercial and business land uses, can significantly boost railway patronage. This can be explained by the fact that these areas act as either trip origination or destination locations for a multitude of trips. The research findings also proved that road network characteristics and connectivity with bus services could have significant effects on attracting more rail users by providing easy access and convenient transfer to rail services. In this regard, promoting high levels of land use and providing a pedestrian friendly environment around station areas (so-called Transit Oriented Development (TOD)), is viewed as a promising planning tool to boost transit usage, discourage car-based travel and mitigate traffic congestion and emissions. The built environment of TOD around rail stations and its effect on promoting rail usage and resolving traffic problems is widely discussed in the literature. Gard (2007), for example, using travel behaviour data for several rail transit lines in California found that, compared with conventional development, TOD increases transit usage by 2-5 times and reduces car trip generation by between 8% and 32%. In another study, Garrett and Castelazo (2004) found that traffic congestion growth reduced in some US cities after introducing a light rail service. In Baltimore the annual average congestion index declined from 2.8% to 1.5% after the light rail service began. In Sacramento, the index plummeted from 4.5% before light rail to only 2.2% afterwards. As a result, TOD concepts have been implemented in many countries, including the United Kingdom, the Netherlands, United States, Hong Kong, Taiwan and China in order to tackle their urban sprawl problems, especially over-reliance on private cars and associated traffic congestion problems (Cervero and Murakami, 2008, Lin and Shin, 2008, Sung and Oh, 2011).

3.2.3.4 Economic Growth

This refers to the ability of a rail transit system to accommodate a community's economic goals, such as increases in land use development, investment, employment, land value, business activity and economic productivity. Establishing a rail system has a strong effect on economic growth by improving accessibility and liveability in the areas surrounding the system, particularly around the station itself. Stimulating economic development and injecting vitality into declining areas have therefore been the major, and sometimes the most important, objectives of building new rail systems or expanding existing ones in many cities around the world. Examples include construction of the rail system in London Docklands, UK, and in Memphis, USA (Mackett and Edwards, 1998).

Research over the last few decades has explored how investment in a rail transit system promotes economic growth at the local and regional levels. A case study in Minnesota, United States, showed a significant increase in housing development and property values after the construction of the Hiawatha rail line, at the proximity of the line and around stations. The amount of new housing construction occurring next to the Hiawatha rail line was 183% more than expected. The average property value of a single family and a multifamily home in a station area (half mile radius from a station) also increased by more than $5,000 and $15,500 respectively. This resulted in a total increase of $47.1 million in residential property within only three years of the line opening in 2004 (CTS, 2009). As Smith et al. (2015) and Cervero and Murakami (2008) have argued, the construction of a rail transit system could often be partly or totally funded through the property value increases it provides. In addition, these results confirm that planning a rail system must not only be linked to transport improvements but also viewed in the broader context of the socio-economic improvements within the potential served area.

3.2.3.5 Environmental Impacts

These refer to the degree to which a rail system disturbs the environment and encounters social complications due to its intrusion into sensitive areas. Rail transit systems have a number of environmental benefits, not just from cutting down on tailpipe emissions but also from preserving green spaces and achieving energy conservation (Chester and Horvath, 2008, Cervero et al., 2004, Litman, 2015a). Nevertheless, as the system is planned, a full range of community impacts must be considered so that environmental and social issues do not emerge. Intrusion of a rail system into areas that are expected to encounter these issues is thus crucial to be considered in determining the locations of rail stations and the associated network of lines connecting them.

3.2.3.6 System Construction Cost

This refers to the capital costs required to construct stations and the sets of lines connecting them, which includes provision for ensuring right-of-way, performing earthworks, building structures (e.g., bridges, tunnels and stations), laying rail tracks, fastening rails and other various items. There are two main reasons for categorising this element under the list of community requirements. First, the construction cost of such infrastructure systems is largely covered by public resources (various sectors of government) and therefore it has an impact on the community. Second, involving this element in the lists of the requirements of other stakeholders, i.e., passengers' and operators' requirements, hinder the impact of the investment cost on the benefit of the system in terms of the passenger service quality and operator productivity. That is, considering this element within the list of community requirements makes it easier to compare alternative solutions based on the degree to which each meets the respective passenger and operator requirements given a particular construction cost. Furthermore, this makes the trade-off between the capital cost and both service quality and operational productivity easier, and thus the decision making process less complex. For example, if two alternative solutions provide similar levels of service quality for passengers and similar operational costs, the community would tend to prefer the one with the lower capital cost. In general, however, higher investment in building rail systems would deliver either higher service quality for passengers or lower operational costs for operators, or a combination of the two. The decision must, therefore, be based on an evaluation of whether the additional passenger benefits, operational cost savings and other positive effects are worth the additional investment cost.

3.3 Level II – Translation of Rail Transit System Requirements

At level II, the requirement sets of the three interested stakeholders: passengers, operators and the community identified in section 3.2 are translated into effective criteria that can be used to evaluate various alternative solutions. The translation of the requirements into sets of criteria is performed based on a comprehensive literature review of rail system planning and the factors upon which the identified stakeholder's requirements (i.e., sub-objectives) depend. The consideration of such factors will help to attain the stakeholders' main objectives.

3.3.1 Passenger Requirement Criteria

The basic objective of passengers in utilizing rail is to arrive at a desired destination quickly, comfortably, safely and at a reasonable price. In the context of rail system planning, the proximity of rail stations to passengers' origins and destinations, and the degree to which they integrate with other transport networks, along with the travel speed along the line network can be considered to be the largest component of utility to the passengers' requirements. These requirements are, however, complex and interlinked, which can make the decision-making process more complex if they are not clearly distinguished from each other. There is a need, therefore, to identify the factors upon which they depend and determine how they interact in order to provide a better insight into the planning process and thus make the decision process less complex and reliable.

3.3.1.1 Station Availability Criteria

As defined in section 3.2.1.1, this is the distance of rail stations from passengers' origin and destination locations. Based on this definition, the key determinant of this requirement is the access / egress distance to/from rail stations. The importance of this element, however, varies with the transport mode used to arrive to, and depart from the rail stations. It holds more significance when walking is the dominant access mode than when the private car or bus is the mode used to connect to rail stations. Previous research has examined how far railway passengers are willing to walk to rail stations and how the distance to and from rail stations influence their access mode decisions. For example, in a case study on the Bay Area Rapid Transit (BART) system in California, Cervero (2001) examined the relationship between rail transit access mode and distance. The results showed that at a distance of 1 km or less the dominant mode of access were walking; for distances between 1 km and 1.6 km bus transit replaced walking as a dominant mode for connecting to the rail stations; beyond 1.6 km the dominant mode of access was the car. Kim et al. (2007), in an empirical study on MetroLink passengers in the St. Louis metropolitan area in the US found a similar pattern. The average distance to stations from passengers' homes was 0.76 km, 5.95 km and 8.85 km for passengers who walked, used buses and private cars to access the stations, respectively. These findings confirm that the sensitivity of potential passengers to walking to rail stations is higher compared to the use of the bus and car, as well as the dominance of walking to stations when the access distance is 1 km or less. This is because, compared to walking, the willingness of potential passengers to use buses or cars to access rail stations is more likely to be influenced by the quality of the rail service, the trip length and the travel conditions of these alternative transport modes.

As discussed in both the operator and community requirement sections the willingness of people to use the rail system decreases significantly beyond walking distance to rail stations. Furthermore, the large majority of land development and economic growth generally occurs within this distance. Therefore, , this thesis uses walking distance to/from rail stations as a dimension to measure the station availability requirement. This assumption is also consistent with the recommendation by Vuchic (2005) for rail transit system planning. He reported that, for planning purposes, it can be considered that the majority of potential passengers whose origin or destination is within 400 m (5 minutes) walking distance will use the system. From this distance and 800 m (10 minutes) the number of potential passengers gradually decreases virtually to zero. The definition of walking distance to rail stations, however, depends largely on whether the system is to be planned for an urban area or a suburban area, the level of service quality at the rail station (e.g. network connectivity, area coverage and service facility and frequency), and the environmental characteristics of a served area under consideration. It therefore, differs from one place to another and thus there is no single standard value of walking distance to rail stations for planning purposes. Most studies in the US used a range between 400 and 800 m (¼ and ½ mile) distance from rail transit stations (O'Sullivan and Morrall, 1996, Evans et al., 2007, Cervero et al., 2004, CTS, 2009). A number of studies in Europe have used a 700 m walking distance boundary from rail transit stations (Zemp et al., 2011, Bertolini, 1999, Reusser et al., 2008) while some Australian studies suggest a boundary of 800 m as a walking distance from rail transit stations (Kamruzzaman et al., 2014). For planning purpose, it can therefore, be concluded that a walking distance of between 400 to 800 m of a station is the most effective range for satisfying passenger needs.

3.3.1.2 Integration with other Transport Mode Criteria

Good integration between rail services and other transport mode services is one of the important goals that both passengers and operators are directly interested in. For passengers , good integration is not only vital so as to connect to rail stations easily and conveniently, but also for reducing trip travel time, as discussed in section 3.2.1.3. As for operators, successful integration plays a key role in increasing the system patronage and associated revenue gains, as explained in section 3.2.2.5. For the community, good integration is an important factor in attracting people to utilize the rail system which tends to reduce dependence on the car, and thus congestion and environmental pollution. Successful integration is therefore, also an important goal for the community.

The integration between rail and other transport modes can be considered as a factor with two main elements: the quantity and quality of the transport infrastructures, such as airport terminals, bus stops, roads and car park facilities, clustered around rail stations; and the level and quality of the coordination between the offered rail services and other transport mode services, particularly bus services. The latter largely depends on the degree to which the frequency and schedule of both the offered train service and bus service are coordinated by the relevant operating company(s). In other words, how the train and bus schedules are coordinated to ensure that a designated bus arrives at a particular bus stop linked to a rail station earlier than the train departure time, taking into account the necessary transfer walking time so that the passengers can successfully access the rail system and complete the reminder of their trips and vice versa. For planning purposes, therefore, the former element can be considered to be the most important dimension of a successful integration. Basically, the larger the number of car parks, bus stops at airport terminals and ports surrounding rail systems, particularly station nodes, the better the connectivity would be.

Extensive availability of car parks or bus stops in the proximity of station areas, for example, can provide passengers with comfortable transfers and reduce out of train travel time (walking, waiting and transferring). These improvements can, in turn, help the rail system operate more efficiently and can make it a more attractive travel option and thus eventually, contribute to increases in rail passengers (Loo et al., 2010). In this regard, many studies have used the number of bus stops, routes and car parking facilities clustered around rail station areas to measure the level and quality of integration between rail and other transport modes. For example, Loo et al. (2010) used the number of bus stops , bus lines and bus headways as variables to represent integration between rail and other transport modes, specifically bus transit services, while examining the TOD planning factors that affect increases in rail transit usage in the city of Seoul, Korea. The results of their analysis, however, showed that a greater number of bus stops within rail station areas, regardless of the number of total bus routes, led to increases not only in rail transit usage but also bus transit usage due to transfers between rail and bus transit services. This indicates that the number and location of bus stops within rail station areas can be a critical planning factor to enhance intermodality among public transport modes and thus create a more transit-oriented city. In another example, Brons et al. (2009) evaluated the impact on passengers of the level and quality of services to rail stations. The overall satisfaction of passengers with rail journeys included four variables representing he quality of the accessibility to the rail stations, connections with public transport, guarded bicycle parking, unguarded bicycle parking, and car parking capacity. The results showed a significant correlation between passengers' satisfaction with each of these variables and their overall satisfaction with the rail journey, except for the satisfaction with the guarded bicycle parking facilities. Connections between the train and public transport (bus/tram/metro) were the most significant features of connectivity to the rail stations, followed by car park capacity and unguarded bicycle parking facilities.

3.3.1.3 Travel Speed Criteria

Train travel speed is one of the most important characteristics of rail transit systems that directly influences both the passengers' travel time and mode choice. Although travel speed can be incorporated into the planning platform directly, it is essential to consider those factors that can affect travel speed to guarantee that the system can operate at or close to its designed speed. The distance between stations and the geometry and gradient of the line network are among the factors that limit train travel speed. A shorter distance between stations results in a decrease in travel speed due to the trains having to stop frequently at stations, which hinders them from reaching their maximum running speed. On the other hand, increasing station spacing tends to increase the access distance to rail stations which results in increased passenger travel time and thereby, reduces their overall satisfaction with a rail journey. This in turn affects the operator productivity level and associated revenue gains.

Reducing the distance between the rail stations leads to increased system usage, due to a decreased passenger travel time, which plays a key role in increasing operators' productivity levels. Rail system planners therefore, face a dilemma in balancing travel speed and station availability requirements. To achieve an efficient rail system it is, therefore, essential not to underestimate the complex correlation of these two requirements on the service quality and operational costs of the system when making a trade-off between them. In addition, the train travel speed is strongly linked to the geometric features of the line alignments connecting stations, such as curvature, gradient and the length of the line alignments. Increasing the tightness of horizontal curves and the steepness of gradients decreases the operating speed of trains on the rail line network. Accordingly, in addition to taking into account travel speed in the evaluation process, this thesis also considers both the station spacing and geometry of line alignments connecting the stations so as to achieve a high service quality and operationally cost effective rail system.

3.3.2 Operator Requirement Criteria

The main operator requirements that must be accounted for when deciding where to place rail stations and how to connect them in a network of lines to obtain the final system configuration are: station availability, population coverage, passenger attraction, travel speed and integration with other transport modes, as described through section 3.2.2. The dominant goal of operators is to attract as many passengers as possible so as to allow the system to operate as economically as possible and thus increase revenue gains. Operators, therefore, devote efforts to satisfying passengers' basic requirements. This means that some of the operators' requirements in practice coincide with the passenger requirements, such as station availability, travel speed and integration with other transport modes. Although these requirements are viewed differently from both the passenger and operator perspective, they are influenced by similar factors. To avoid duplication, therefore, this section only focuses on population coverage and passenger attraction elements.

3.3.2.1 Population Coverage Criteria

As defined in section 3.2.2.2 and 3.2.3.2, population coverage refers to the number of people who live within usual walking distance of a rail transit station. Maximizing population coverage is a very important goal in planning a rail system from both the operator and community perspectives. If the system does not attract enough passengers the operator will suffer losses and the community will not benefit from mitigation in dependence on private cars and associated traffic and pollution problems. Population coverage can be considered to be a function of two main elements: the population density around rail station sites and the station density along the line network, i.e., the number of stations per unit distance along each line in the network.

Research exploring the planning factors that are expected to contribute to increases in railway patronage indicates that increased population density within walking distance of station areas is associated with higher usage of the system (Cervero, 1993, Cervero and Kockelman, 1997, Cervero et al., 2004, Evans et al., 2007, Frank and Pivo, 1994, Yoh et al., 2003). This can be attributed to two main factors. Firstly, a high population density is associated with high travel demand and therefore, tends to increase the use of the rail system. Secondly, a high population density is often associated with restraints for car use, such as lack of parking space, high traffic volume and low travel speed, and thus reduces the comparative disadvantages of public transport against private cars, making the former a more attractive option for travel (Scheiner, 2010). Population density is widely considered to be the most important predictor of usage rates in rail system planning and has almost become a standard scale for measuring the system coverage. Furthermore, a high density of stations in a rail network results in an increase in the total system coverage. However, this is associated with high operational costs because of increased maintenance costs due to frequent stops at stations (Van Nes and Bovy, 2000, Samanta and Jha, 2011). This is in addition to the high cost associated with constructing the system infrastructure, which might not be practical, particularly due to budget constraints. Accordingly, this thesis incorporates population density and station density into the planning platform in order to measure the population coverage of the proposed rail system while considering the associated large investment and operational costs due to high station density.

3.3.2.2 Passenger Attraction Criteria

As discussed in sections 3.2.2.3 and 3.2.3.3, the dominant requirement of both operators and the community in planning a rail system is to attract as many passengers as possible. The number of people who use the rail system is a function not only of the potential demand, but also of the degree to which it meets the passenger requirements defined through section 3.2.1, and coordinates with land use patterns to facilitate various social activities, such as marketing, shopping and tourism. The coordination of rail transit systems with land use patterns, particularly high-density commercial land use can be viewed as a promising tool for coaxing people out of their private cars and into trains by providing a good walking environment, mobility and accessibility to activities surrounding rail stations. In the UK, for example, among the main factors behind the success of the Manchester Metrolink light rail system in attracting more car users than predicted and allowing fare revenue to exceed operating costs was its penetration into the city centre (Babalik, 2000, Knowles, 1996). The city centre of Manchester is identified as one of the main locations for business and retail development. This finding reinforces the conclusions of other studies examining factors that influence the attractiveness of rail transit systems and the degree to which this contributes to encouraging people to switch from using private car to rail (Cervero, 1993, Cervero et al., 2004, Evans et al., 2007, Cervero and Murakami, 2008, Loo et al., 2010, Sung and Oh, 2011). These research findings confirm that promoting high residential, commercial and business density land uses and providing pedestrian-friendly environments around station areas can significantly boost railway patronage.

In analyses of rail usage among land use developments near railway stations in Washington D.C., JHK and Associates (1987) estimated that 200,000 square feet (about 18500 square metres) of additional office building in the downtown area would generate nearly 300,000 additional transit trips per year, which would yield approximately $500,000 in transit revenue. Also, a similar level of building near suburban rail station areas would being over $200,000 in additional transit revenues annually. In a similar study using the 2000 US census data, Cervero et al. (2004) found that every 100,000 square feet (9290 square metres) of additional office and retail floor space near an Arlington County Metrorail station increased the average daily boarding and alighting patronage at that station by 50. In a more recent study, Loo et al. (2010) provided further support to the notion that land use and street network characteristics around rail stations stimulate railway patronage and reduce dependence on cars. Their results showed that when the commercial floor area around a rail station in Hong Kong increased by 10,000 square metres the average weekday railway patronage would tend to increase by 100. For planning purposes therefore, it can be considered that the intensity of commercial land use, including office complexes, retail areas and recreational centres, around station areas is an important indicator for measuring the attractiveness of the rail system.

3.3.3 Community Requirement Criteria

As discussed in section 3.2.3, the main community requirements are economic growth and environmental improvements. However, they are strongly influenced by line network configuration and associated station locations, and thus are also linked with passenger and operator requirements. This large overlap between the community requirements and the other two stakeholders' requirements is simply because the community aims to realise a rail system that can offer a suitable service to the entire population of the served area, including passengers and operators. Although these overlapping requirements take different forms according to the particular perspective of the three interested stakeholders, as discussed in section 3.2, they have identical influences. To avoid duplication, therefore, this section focuses on the economic growth and environmental impact elements.

3.3.3.1 Economic Growth Criteria

Supporting economic development is the most important long-term goal for rail transit system planning from the community perspective. This includes improvements to land use development, investment, employment, land value, business activity and economic productivity, as discussed in section 3.2.3.4. Satisfying these requirements largely depends on the degree to which the planned rail system is integrated into important activity centres, existing development projects, and deprived areas that need regeneration. The other relevant factors are the degree to which the local municipalities or urban planning agencies encourage investment and redevelopment in and around the system. This demands good coordination between transport and urban planning policies. When transport planning policies, for example, call for making rail system plans, compatibility with urban development strategies is very important. At the same time, planning policies should also support sufficient flexibility to allow urban plans to adapt to rail system investment and development. There is considerable empirical evidence to show how coordination between transport and urban policies has contributed to the success of rail systems in attaining their economic objectives. For example, the most important factor for the great success of the Vancouver SkyTrain is attributed to the strong support of municipalities for integrating the SkyTrain into the local plan and projects (Babalik, 2000). During the construction of the system, municipalities, in addition to redeveloping old industrial areas around the system, adapted their local plans to the system by implementing joint development schemes, introducing developments to private developers and relocating some government buildings at SkyTrain stations. In contrast, poor coordination of policy between the regeneration and transport planning agencies were the major reason for the limited input of the Sheffield (UK) super tram in revitalizing and redeveloping the deprived areas along the tram route (Babalik, 2000).

Various techniques and models can be used to measure the impact of rail transit systems on economic developments, such as cost-benefit analysis, transportation-land use models, economic forecasting models and input-output tables (Litman, 2010, Chatman et al., 2012, Hazledine et al., 2013, Litman, 2015b). Since the evaluation of the impact of the rail system on economic development lies beyond the scope of this research it is assumed that those areas that have potential to bring maximum possible economic growth to the served area have been identified prior to the evaluation process. It is also assumed that those areas, including deprived areas that need regeneration, undeveloped land parcels that have strong potential for economic growth, and existing development projects, have been identified by the related transport and urban planning agencies. To secure improved economic development, therefore, the identified areas are incorporated into the decision-making framework so that they can be taken into account when deciding where to place stations and how to connect them in a network of lines.

3.3.3.2 Environmental impact criteria

Improving environmental conditions can also be considered to be one of the primary goals of planning the rail transit system from the community perspective. Despite the significant positive role that a rail transit system can play in reducing pollution emissions and energy consumption it may still affect the environment negatively. Such environmental impacts are often considered to be important issues in building rail transit system projects (Lai, 2012). Consideration in the planning stage of the wider environmental effects of rail transit construction projects is therefore essential so that environmental and social complications do not occur. Such effects mainly include intrusion into environmentally sensitive areas located along the proposed rail lines or near stations. Such environmentally sensitive areas may include parks, sites of architectural and historical interest, woodlands, forests, wetlands and sites of scientific interest. The extent of impact on these areas, however, depends on the type of the proposed rail system, i.e., whether the system is to be built above ground or underground. Parks, for example, should be identified as environmentally sensitive areas if the system is to be built above ground, whereas for underground rail systems these same areas can be identified as unlikely to be liable to environmental consequences. These elements should therefore, be considered as criteria to satisfy the environmental requirements.

3.3.3.3 System Construction Cost Criteria

As discussed in section 3.2.3.6, system construction costs are the capital costs required for constructing stations and the line network connecting them. These elements depend not only on the system size and construction type (i.e., whether it is constructed as an underground or an above ground system), but also on the locations of stations and associated line network. For planning purposes, the system construction cost items can be classified into two main categories; location-dependent costs and size-dependant costs. The first category involves those construction items which change significantly with different locations for the same station size, total line network length and construction type. The second category includes those construction items that do no change with different locations for the same station size, total line network length and construction type. For the stations, associated costs therefore for building and equipping stations with the required facilities can be listed under the size-dependent construction cost category while the land acquisition and earthwork costs can be classified as location-dependent construction costs. As for the rail line set, the size-dependent costs involve only costs for laying track and fastening rails while the location-dependent costs comprise the rest of the associated rail line cost items (i.e. land acquisition costs, earthwork costs, and bridge and tunnel costs). Table 3.1 summarizes the dominant objectives and corresponding requirements (sub-objectives) of the three interested stakeholders defined in section 3.2 and presents the dominant dimension of each requirement discussed in section 3.3.

Table 3.1: Objective and Corresponding Requirements and Criteria of the Three Interested Stakeholders.

Stakeholders	Objectives	Requirement sub-objectives	Planning related requirement sub-objectives	Requirement criteria
Passenger	Arrive at a desired destination within as short a time as possible, comfortably, safely and at a reasonable price.	• station availability • travel speed • comfort • Punctuality • integration with other transport modes • safety at stations and on board • fare cost.	• station availability • travel speed • integration with other transport modes	• walking distance to stations • speed, station spacing and line alignment geometry • number of airport terminals, bus stops and car parking facilities around station areas
Operator	Maximize the rail system revenue gains by attracting the maximum possible number of passengers, and at the same time minimizing operational costs.	• station availability • travel speed • integration with other transport modes • population coverage • passenger attraction • capacity • operation reliability and safety • fare cost.	• station availability • travel speed • integration with other transport modes • Population coverage • passenger attraction	• walking distance to stations • speed, station spacing and line alignment geometry • number of airport terminals, bus stops and car parking facilities around station areas • population density around station areas and station density • commercial land use areas around station areas
Community	Provide a rail system that is easily available to all residents with a high service quality to attract many people, and at the same time reduce traffic congestion and enhance economic activities with minimal environmental impacts and investment costs	• station availability • Population coverage • passenger attraction • economic growth • environmental impacts • construction cost.	• station availability • Population coverage • passenger attraction • economic growth • environmental impacts • construction cost	• walking distance to stations • population density around station areas • commercial land uses areas around station areas • deprived areas or undeveloped land areas with strong potential for economic growth around station • areas of existing development projects around station • Environmentally sensitive areas • Associated costs for building stations and lines

3.4 Level III – Generation of Alternative Solutions and Optimization

This level formulates mathematical functions of the three interested stakeholders' requirement criteria, identified at level II, and incorporates these into a single planning platform within the context of an integrated optimization model to achieve a rail transit system that best fits the desired objectives defined at level I. The proposed model aims to directly incorporate the determination of station locations and network of lines, linking them into a single optimization process.

Since it would be very difficult to check the feasibility of every point in the study area as a potential station location, the model is broken down into two stages, as shown in figure 3.2. The first stage screens the study area to identify a pool of feasible station locations based on a comprehensive consideration of the various requirement criteria identified at level II. The second stage selects a subset of feasible stations from the candidate station pool identified in the first stage and generates a network of lines to link these within the context of an integrated optimization process so as to obtain the final system configuration. The optimization framework is designed to accommodate complex correlations and interactions of the three stakeholders' requirements and make trade-offs between them in order to achieve a reliable and cost effective system. These two stages of the evaluation framework are discussed in detail in sections 3.4.1 and 3.4.2.

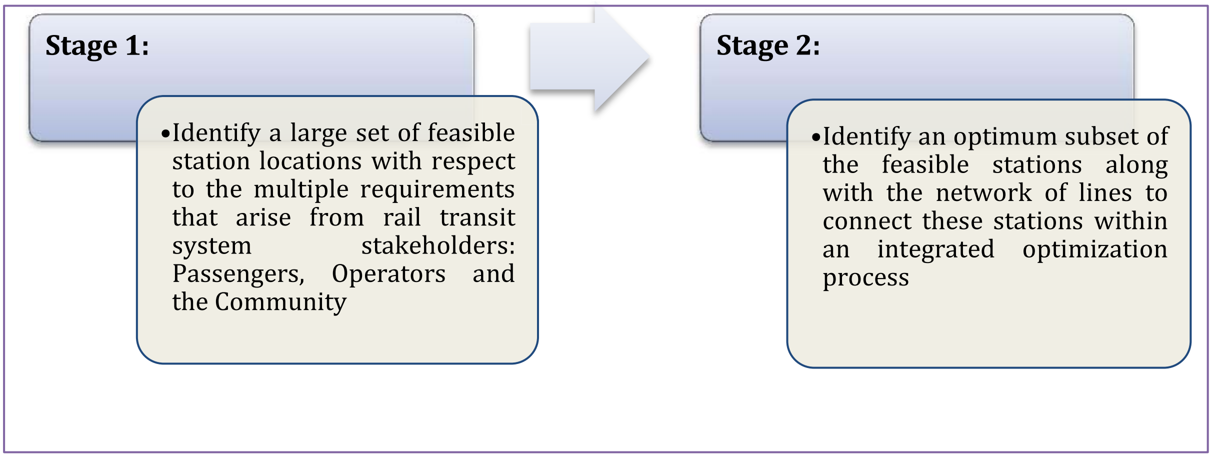


Figure 3.2: Stages of the Proposed Integrated Optimization Model

3.4.1 Stage 1: Determination of Potential Rail Transit Station Locations

This stage measures the feasibility of station sites with respect to the various requirement criteria identified at level II and designates a pool of feasible station locations that best fit the desired objectives of the three interconnected decision-making parties defined at level I: passengers, operators, and the community. However, some of the requirement criteria, such as train travel speed along the line network and station density, can be imposed into the evaluation framework only once a network of lines linking stations has been generated. These criteria are not accounted for at this stage. This by no means excludes them from the planning platform as they will be integrated into the evaluation framework when the network of lines linking stations has been generated. In addition, because all three of the interested stakeholders are united in requiring a rail transit system that can offer a suitable service, there are large overlaps between their corresponding requirements, as described in the preceding sections. The overlapping requirements include maximum operating speed for passenger convenience and operational efficiency and short access/egress distance to/from rail stations for passenger convenience and maximum population coverage for operator and community benefits. To avoid duplication and unnecessarily complex analysis, the coincident requirement criteria are considered once in the evaluation process, as shown in table 3.2.

Table 3.2: Requirement Criteria of Station Location Evaluations

Stakeholders	Rail transit planning requirements	Requirement criteria	Station location evaluation criteria
Passenger	• station availability • travel speed • integration with other transport modes	• walking distance to stations • speed, station spacing and line alignment geometry • number of airport terminals, bus stops and car parking facilities around station areas	• walking distance to stations • number of airport terminals, bus stops and car parking facilities around station areas
Operator	• station availability • travel speed • integration with other transport modes • Population coverage • passenger attraction	• walking distance to stations • speed, station spacing and line alignment geometry • number of airport terminals, bus stops, car parking facilities around station areas • population density around station areas and station density • commercial land uses areas around station areas	• population density around station areas • commercial land uses areas around station areas
Community	• station availability • Population coverage • passenger attraction • economic growth • environmental impacts • construction cost	• walking distance to stations • population density around station areas and station density • commercial land use areas around station areas • deprived areas or undeveloped land areas, which have strong potential for economic growth, around station • areas of existing development projects around station • Environmentally sensitive areas • Associated costs for building stations and lines linking the stations.	• deprived areas or undeveloped land areas, which have strong potential for economic growth, around station • areas of existing development projects around station • Environmentally sensitive areas • Associated costs for building stations.

Using a geographic information system (GIS), this stage evaluates the feasibility of station locations with respect to the requirement criteria presented in table 3.2. GIS tools have been widely deployed in solving transport design problems due to their ability to retrieve large amounts of real data such as land value, land use patterns, topographic information and existing street networks in addition to performing spatial analysis. The following five main steps are applied to formulate mathematical functions for the identified requirement criteria, and to perform evaluative and comparative analysis in order to generate feasible locations for rail transit stations within the context of a GIS-based algorithm.

Step 1: divides the study area into grids (G_i) and creates a GIS layer (Ω_s) for stations. It is assumed that each of the generated grids represents a potential location for a rail station. The size of each grid, therefore, should represent the typical size of a rail station.

Step 2: evaluates the locations of the station grids (Gi) with respect to the requirement criteria of the three stakeholders, illustrated in table 3.2, using their predefined respective datasets in ArcGIS. For instance, the evaluation of passenger requirements consists of counting the total number of airport terminals, bus stops and car parking facilities clustered within a defined walking distance to stations.

Step 3: finds all station grids that intersect with environmentally sensitive areas and excludes them from the search space of potential station locations. This is done by laying the generated station grid layer (Ω_s) over the environmentally sensitive areas layers (Ω_e) to generate the feasible grid layer for stations(Ω_f). The environmentally sensitive areas include historic buildings, national parks, woodlands, forests, rivers and sites of scientific interest.

$$\Omega_f = \Omega_s \cap \overline{\Omega_e} \tag{3.1}$$

Step 4: identifies all the grids within the feasible station layer (Ω_f) that satisfy the predefined conditions set by the threshold values of the requirement criteria and assigns them with integer values, which represent the weight of the satisfied criteria, as follows:

$$\forall G_i \in \Omega_f \rightarrow S^i = \{1, \text{ if } (\sum AT \,|D_w) \geq N_{AT} \tag{3.2}$$

Where: AT is the number of airport terminals (if any) within the defined walking distance, D_w is the defined walking distance and N_{AT} is the pre-specified threshold number of airport terminals (if any) within D_w.

$$\forall G_i \in \Omega_f \rightarrow S^i = \{1, \text{ if } (\sum RS \,|D_w) \geq N_{RS} \tag{3.3}$$

Where: RS is the number of existing railway stations (if any) within D_w and N_{RS} is the pre-specified threshold number of existing railway stations (if any) within D_w.

$$\forall G_i \in \Omega_f \rightarrow S^i = \{1, \text{ if } (\sum CP \,|D_w) \geq N_{CP} \tag{3.4}$$

Where: CP is the number of car parks within D_w, N_{CP} is pre-specified threshold number of car parks within D_w .

$$\forall \mathrm{G_i} \in \Omega_\mathrm{f} \rightarrow \mathrm{S^i} = \{1, \text{ if } (\sum \mathrm{BS} \,|\mathrm{D_w}) \geq \mathrm{N_{BS}} \tag{3.5}$$

Where: BS is the number of Bus Stops within $\mathrm{D_w}$ and $\mathrm{N_{BS}}$ is pre-specified threshold number of bus stops within $\mathrm{D_w}$..

$$\forall G_i \in \Omega_f \rightarrow S^i = \{1, \; if \; (\overline{PD}|D_w) \geq V_{PD} \tag{3.6}$$

Where: PD is the average population density within $\mathrm{D_w}$ and $\mathrm{V_{PD}}$ is pre-specified threshold value of average population density within $\mathrm{D_w}$.

$$\forall \mathrm{G_i} \in \Omega_\mathrm{f} \rightarrow \mathrm{S^i} = \{1, \text{ if } (\sum \overline{CA} \,|\mathrm{D_w}) \geq \mathrm{V_{CA}} \tag{3.7}$$

Where: $\overline{CA}$ is the average commercial land use Areas within $\mathrm{D_w}$, $\mathrm{V_{CA}}$ is the pre-specified threshold value of average commercial land use area within $\mathrm{D_w}$. commercial land uses include office complexes, retail areas and recreational centres.

$$\forall \mathrm{G_i} \in \Omega_\mathrm{f} \rightarrow \mathrm{S^i} = \{1, \text{ if } (\sum \overline{DA} \,|\mathrm{D_w}) \geq \mathrm{V_{DA}} \text{ OR if } (\sum \overline{DP} \,|\mathrm{D_w}) \geq \mathrm{V_{DP}} \tag{3.8}$$

Where: $\overline{DA}$ is the average deprived areas that need regeneration or undeveloped land areas that have a strong potential for future economic developments within $\mathrm{D_w}$, $\mathrm{V_{DA}}$ is the pre-specified threshold value of average area of deprived /undeveloped land parcels within $\mathrm{D_w}$, $\overline{DP}$ is the average area of existing development projects within $\mathrm{D_w}$ and $\mathrm{V_{DP}}$ is the pre-specified threshold value of average area of development projects within $\mathrm{D_w}$.

$$\forall \mathrm{G_i} \in \Omega_\mathrm{f} \rightarrow \mathrm{S^i} = \{1, \text{if } (\overline{\mathrm{LV}}|\mathrm{D_w}) \leq \mathrm{V_{LV}} \tag{3.9}$$

Where: $\overline{\mathrm{LV}}$is the average Land Value within $\mathrm{D_w}$ and $\mathrm{V_{LV}}$ is the pre-specified threshold value of the maximum allowed land acquisition value (unit cost per unit area) for stations.

The weight (integer values) attributed to each condition represents the degree of importance of the corresponding requirement criteria; the higher the weight, the higher the degree of importance. For simplicity, initially equal weights of 1 are assigned to all the requirement criteria in equations 3.2 to 3.9, assuming equal importance of the requirements. The algorithm can accommodate different weights for different requirement criteria, for cases where the prioritization of a particular requirement is necessary.

Step 5: aggregates the weights of the corresponding requirement criteria of the three stakeholders that are satisfied by station grids and designates all grids that meet the pre-defined satisfaction level.

In this thesis, the satisfaction level is defined as the minimum aggregate weight of the three interested stakeholders' requirement criteria to be met by the potential station locations. The higher the satisfaction level, therefore, the larger the number of requirement criteria that need to be satisfied by each potential station grid. Due to the conflicting nature of some of the requirements, such as low land value area and high density commercial land use areas, it is very difficult to find a sufficient number of locations that can satisfy all the requirement criteria. The satisfaction level is therefore, implemented in the proposed algorithm to enable rail transit planners to make trade-offs between the stakeholder requirements and thus to obtain a practical number of feasible station locations for the optimization process to choose from. Chapter 4 provides the details for implementing these steps within the GIS-based algorithm.

3.4.2 Stage 2: Optimization of Station Locations and Associated Line Network

Once the set of feasible station locations is designated using the GIS-based algorithm, the second stage seeks the final configuration of the proposed system within the context of an integrated optimization process. It simultaneously selects the best subset from the potential station set identified in the first stage along with generating the best line network to link the selected stations. In addition to the comprehensive consideration of various local factors, including travel demand patterns, land use patterns, topography and soil type, the proposed optimization algorithm takes into account the multiple requirements of the three interested stakeholders' objectives identified in the preceding sections. Due to the complexities of the rail transit system planning requirements, which involve a non-differentiable, nonlinear, discontinuous structure, it is very difficult to model them with simple mathematical functions, as discussed in chapter 2. Therefore, a heuristic search method based on the genetic algorithm (GA) is designed in this stage to efficiently search through the decision variables of the three interconnected stakeholder requirements and find a solution that best fits their corresponding desired objectives. In addition, GIS is embedded within the search method, which can efficiently retrieve and analyse large amounts of real data such as land value, land use patterns, topographic information and soil type and thus makes the GA search algorithm more efficient in finding a robust solution to the problem. The following two sub-sections present the framework of the optimization algorithm and elaborate the formulation of the three stakeholders' requirements into a set of objective functions and constraints.

3.4.2.1 The Optimization Objective Functions

As discussed in section 3.3, some requirements of the three interested stakeholders overlap, whereas others contradict. The most mutually contradictory requirements are minimum investment cost for community benefits versus maximum population coverage for passenger service quality and operator productivity. Generally, locating stations in well-developed areas (i.e. high density residential or commercial land use areas) and increasing station density along the system line network are viewed as the most effective planning strategies to maximize railway patronage, as explained in section 3.3. These strategies are, however, associated with high investment cost since areas with high density residential or commercial land use almost always have high land values, which makes building stations in these areas costly. Increasing the station density is also strongly associated with high investment cost for the same reason. Therefore, it is difficult to reconcile the maximization of rail transit system patronage with minimal investment costs and it is therefore important to arrive at a balance between these requirements in order to achieve a reliable and cost-effective rail system. Consequently, the proposed optimization algorithm is designed to resolve this trade-off between adequate service quality and benefits for both passengers and community and economically efficient operation and investment cost. The algorithm minimizes the total system cost, which is a function of passenger cost, operational cost and community cost when selecting station locations and generates the line network in between them. The formulation of these objective functions is further elaborated below:

1- Passenger cost**:** this is formulated by calculating the time cost difference between utilizing the train and other transport modes, specifically the bus and car. Assuming the proposed system will offer a service on a daily basis, the annual passenger cost (P_c) will be:

$$P_c = 365 \times \sum_{l=1}^{q} \Big(\sum_t \sum_i \sum_j \big((P_{ij} \times T_{Cttij} - P_{cij} \times T_{Ctcij}) + (P_{ij} \times T_{Cttij} - P_{bij} \times T_{Ctbij}) \big) \quad (3.10)$$

Where:

q is the number of lines in the network,

P_{ij} is the number of railway passengers from zone i to zone j in time period t;

T_{Cttij} is the cost (£) of the total travel time by train from zone i to zone j in time period t, which includes access time to/from the station, waiting time at the station and on-train travel time;

P_{cij} is the number of railway passengers from zone i to zone j in time period t that may switch to car if the rail system is not available;

T_{Ctcij} is the cost (£) of the total travel time by car from zone i to zone j in time period t, which includes in-car travel time and time spent searching for a parking space;

P_{bij} is the number of railway passengers from zone i to zone j in time period t that may switch to bus if the rail system is not available;

T_{Ctbij} is the cost (£) of total travel time by bus from zone i to zone j in time period t, which includes access time to/from the bus stop, waiting time at the bus stop and on-bus travel time.

The details of how these elements are calculated are presented in chapter 5, section 5.4.1.

2- Operator costs: these are formulated to calculate the operation and maintenance cost differences between utilizing rail and other transport modes, specifically car and bus. The operation and maintenance costs of each of these modes is a function of their corresponding travelled distance and potential passengers. Thus the annual operating cost (O_c) is calculated as follows, assuming the rail transit will offer a service on a daily basis as noted earlier:

$$O_c = 365 \times \sum_{l=1}^{q} \left(\sum_t \sum_i \sum_j \left((P_{ij} \times L_{tij} \times M_{t0} - P_{cij} \times L_{cij} \times M_{c0}) + (P_{ij} \times L_{tij} \times M_{t0} - P_{bij} \times L_{bij} \times M_{b0}) \right) \right) \quad (3.11)$$

Where:

L_{tij} is train km travelled distance from station i to station j in time period t;

M_{t0} is unit operational and maintenance cost for the train (£/ passenger-km);

L_{cij} is car km travelled distance from zone i to zone j in time period t;

M_{c0} is unit operational and maintenance cost for car (£/ passenger-km);

L_{bij} is bus km travelled distance from zone i to zone j in time period t and;

M_{b0} is unit operational and maintenance cost for bus (£/ passenger-km).

Details of how these elements are calculated are presented in chapter 5, section 5.4.2.

3- Community costs: these are modelled to comprise the construction costs of the system, and include the capital cost required for constructing stations and the set of lines connecting the stations. As illustrated in section 3.3.3.3, the construction cost items are broken into location-dependent and size-dependent cost categories. The former category involves those construction items that change with different locations of stations and the associated line network, whereas the latter category covers those construction cost items that are proportional to the length of the line network and size of the associated stations. Thus, the location-dependant cost category involves the costs for acquiring land and constructing tunnels and escalator barrels, while the size-dependant category covers the costs associated with building and equipping stations with the required facilities and laying rail track and fastening rails. Thus, the community cost ($\mathrm{C_c}$) is calculated as follows:

$$\mathrm{C_c} = C_{LDep} + C_{SDep} \quad (3.12)$$

Where:

C_{LDep} is the location-dependent cost, which is a linear function of the land acquisition cost and tunnel and escalator barrel construction cost and ;

C_{SDep} is the size-dependent cost, which is a linear function of the cost of building and equipping the stations with the necessary facilities and the cost of laying rail tracks and fastening rails.

The details of how these construction items are calculated are comprehensively demonstrated in chapter 5, section 5.4.3.

It should be noted that some requirement criteria identified in table 3.1, such as commercial land use areas, development projects and the prevalence of bus stops and car parking facilities around station areas are not incorporated into the optimization framework. This is because these requirement criteria are largely influenced by locations of stations and they are, therefore, comprehensively accounted for in the identification of feasible station location candidates. These candidate stations are then used as a pool from which the optimization algorithm selects stations for the final system configuration. Incorporating them into the optimization framework in addition to their diminishing value for the decision-making process makes the analysis more complex without having an impact on the results of the optimization. This indicates the importance of performing the feasibility study for station locations prior to the determination of the final system configuration if the complexity of both the decision-making and analysis process is to be kept to a minimum.

It is also very important to note how the proposed optimization algorithm covers the identified requirements of rail transit system planning and addresses the essential trade-offs between the requirements, particularly between both passenger travel time and rail system usage and construction costs. To deliver adequate service quality for railway passengers, the algorithm tries to either decrease passengers' rail travel time compared to that of the bus and car or increase railway usage, or a combination of the two, through the passenger cost function. It also tries to increase the system revenue gains and to achieve an economically efficient operation of the system through the operator cost function. This function calculates the total operation and maintenance cost reductions that can be achieved by using rail instead of the other motorized transport modes, thus the algorithm tries to increase the railway usage in order to maximise this reduction. Furthermore, through the community cost function, which covers all the construction cost items identified in section 3.3.3.3 for building the system, the algorithm tries to minimize the investment costs. The optimization algorithm therefore, resolves the trade-offs by trying to increase the railway riders to reduce both the passenger and operator costs, without significantly increasing the investment costs. This implies that the algorithm may increase the investment costs to a certain extent to accommodate more railway patronage while on the other hand reducing passenger and operator costs by attracting more car and bus users to switch to rail.

3.4.2.2 The Optimization Constraints

In addition to the above objective functions, a number of constraints are embedded in the optimization framework to further balance the conflicting requirements of the rail transit system and enhance the efficiency of the proposed system in terms of both its operation and functionality. These constraints are:

1- Station density: this is the number of stations per unit distance along each line in the network. As discussed in section 3.3.2.1, station density is a very important determinant of rail transit system coverage. High station density contributes not only to increased railway patronage but also to increased passenger satisfaction with the rail journey by reducing their access/egress distance to/from rail stations. On the other hand, high station density, in addition to its association with high investment costs, may result in a significant increase in the system operating cost due to frequent train stops at the rail stations. Subsequently, station density (Ns) is incorporated into the optimization framework as a constraint to make a balance between adequate service and benefits for the public on the one hand, and economically efficient operation and investment costs on the other. This constraint attempts to increase passenger service quality and operator productivity level by preventing the generation of solutions with a station density lower than a predefined minimum (Ns_{min}). Conversely, it tries to reduce the system investment and operating costs by not allowing the generation of a line network with a station density larger than a predefined maximum station density (Ns_{max}).

$$Ns_{min} \leq Ns \leq Ns_{max} \qquad (3.13)$$

2- Station spacing: this refers to the distance between each two successive stations along the rail line network. Although the station density constraint restrains the number of stations along each line in the network, it does not consider how the stations are spaced along the corresponding lines. To minimize passengers' travel time it is very important to space stations with respect to the distribution of their origins and destinations along the line network. Very short distances between rail stations, however, result in reduced train operating speed and increased operation and maintenance costs due to frequent train stops.

The station spacing constraint is therefore, embedded into the optimization framework to control operational requirements and minimize passenger travel time to an appropriate extent. Accordingly, the distance between any two successive stations (Δs) along the rail line network must be between a predefined minimum (Δs_{min}) and maximum (Δs_{max}).

$$\Delta s_{min} \leq \Delta s \leq \Delta s_{max} \tag{3.14}$$

3- Line network connectivity: connectivity among rail lines is very important for both passenger convenience and attraction. Accordingly, this constraint is embedded into the optimization framework which requires each line in the network to intersect with other lines. To ensure that the rail line network is well interconnected this constraint requires that the number of intersected/ transfer stations along each line in the network (Cn) to be equal or greater than a predefined number of intersected/transfer stations (Cn_{min}) along that particular line in the network.

$$Cn \geq Cn_{min} \tag{3.15}$$

4- Line network overlapping: this refers to overlapped or common sections between any two lines in the network. Typically, the more the lines are interconnected the more overlapped sections are likely to be in the network. As explained in the previous section connectivity between rail lines is a very important factor for a rail transit system to attract and efficiently serve the maximum possible number of passengers. Nevertheless, to achieve a rail system that can offer service to a large majority of people in a served area with an acceptable investment cost, it is important to avoid generation of very large overlapping sections between any two lines in the network. Consequently, an overlap rate constraint is incorporated into the optimization process. This constraint, first, calculates the overlap rate between each two intersected lines in the network by dividing the total length of their corresponding common sections to the total length of the corresponding lines. Then it compares the calculated overlap rate (On) of each intersected line in the network with a predefined maximum overlap rate (On_{max}) to make sure that each intersected line in the network satisfies this requirement.

$$On \leq On_{max} \tag{3.16}$$

5- Line network population coverage: this expresses the ratio of the population who live within walking distance of stations along each line (l) in the network (P_s) to the total population who live within walking distance of a particular line (P_l). This constraint is also embedded into the optimization framework in order to guarantee effective use of the proposed network and its services.

$$P_{cl} = \frac{\sum_{i=1}^{n} P_{si}}{P_l} \geq P_{cmin} \tag{3.17}$$

Where:

P_{cl} is population coverage of line *l* in the network;

n is the number stations that are located on line *l*;

s_i is the associated station of line *l*;

P_{si} is the number of people who live within walking distance of (s_i);

P_l is the number of people who live within walking distance of line (*l*) and;

P_{cmin} is a predefined minimum population coverage.

Obviously, the greater the value of P_{cmin} the stronger the interrelationship between the system and the area it services, thus giving, in practice, better passenger service, higher operator profits and greater community benefits.

Finally, with respect to the objective of minimum total system cost, which is a function of the formulated passenger, operator and community objectives, while satisfying the station and line network constraint sets, the GA-based algorithm developed seeks the search space for the optimal solution of the following problem:

$$\text{Minimize } (T_c = \varphi p P_c + \varphi_o O_c + \varphi c C_c)$$

$$\text{Subject to } Ns_{min} \leq Ns \leq Ns_{max}, \Delta s_{min} \leq \Delta s \leq \Delta s_{max}$$

$$Cn \geq Cn_{min}, On \leq On_{max}, \text{ and } P_c \geq P_{cmin}$$

Where: T_c is total system cost, and φp , φo and φc are the coefficients of passenger, operator and community costs respectively. The reason for integrating these coefficients into the evaluation process is to incorporate flexibility into the developed algorithm so as to allow rail transit planners to prioritize a particular stakeholder requirement over the others when required.

3.5 Summary

This chapter has proposed an optimization methodology for planning rail transit systems that can effectively account for multiple requirements and constraints arising from different rail transit system stakeholders. The methodology consists of three main levels of analysis and decision-making. Level I captures the various requirements of the three different stakeholders based on a detailed and comprehensive literature review. The analysis and decision-making process at level II includes the translation of the three stakeholders' requirements identified at level I into effective criteria that can be used for evaluating and comparing different alternative solutions. The translation of requirements into sets of criteria is performed based on the factors upon which the identified stakeholder's requirements depend. Level III formulates mathematical functions for the criteria identified at level II, and incorporates them into a single planning platform within the context of an integrated optimization model to achieve a rail transit system that best fits the desired objectives identified at level I.

Level III consists of two main stages. The first stage develops a GIS based algorithm to screen the potential study area to identify a set of feasible station locations with respect to various station location related requirement criteria identified at level II. This stage aims to help rail transit planners to arrive at an initial evaluation of different station locations, which can effectively speed up the planning process and improve decision-making in selecting reliable and cost effective solutions. This is also a crucial step towards the development of an optimization algorithm that can determine optimal station locations and the network of lines connecting them in a single integrated process. The second stage develops a heuristic optimization algorithm, based on GA, to identify concurrently an optimum set of station locations from the pool of feasible stations identified in the first stage, as well as generating the line network connecting them through interaction with the supporting GIS system. The generated network system satisfies constraints in respect to both station location and the configuration of the line network, and achieves a desirable trade-off between the passenger convenience, operator productivity and community benefit related requirements.

Chapter 4 elaborates the development of the GIS-based algorithm for generating a candidate pool of potential stations based on a comprehensive consideration of the various rail transit system planning requirements. It also presents the application of the algorithm to a real world case study, the city of Leicester. Chapter 5 details the development of the optimization algorithm to select the best set of stations from the candidate station pool and concurrently to generate an optimal line network connecting the selected stations. This chapter also details how the optimization algorithm accounts for local factors, incorporates the multiple objectives and constraints that arise from passenger, operator and community requirements and computes the optimal solution. Chapter 6 details the application of the proposed optimization model in a real world case study, City of Leicester (UK), to examine its effectiveness in finding a robust solution in a large region. An extensive numerical study is also included in this chapter to reveal the importance of screening the study area for feasible station locations prior to the optimization process for the performance of the optimization model.

Chapter 4 – Determination of Potential Rail Transit Station Locations

This chapter addresses Level III of the new framework presented in Chapter 3 for planning a rail transit system and presents in detail the development of the GIS based algorithm to determine the feasibility of station sites and identify the candidate pool of potential station locations. This feasibility analysis aims to help planners to determine a set of potential station locations with a comprehensive consideration of the various planning requirements based on systematic evaluation and comparative analysis. This is a crucial step toward the acquisition of a reliable rail system that can have a positive impact on the area it serves by reducing congestion and environmental damage, improving mobility and supporting economic development. Furthermore, the analysis serves to narrow down the search space for the optimization algorithm by excluding unfeasible areas making the evaluation process less complex and the decision-making process more reliable.

Section 4.1 outlines the requirements of the three stakeholders in rail transit system planning; passengers, operators and the community, that were identified in chapter 3 in respect to the determination of feasible station locations. This section also presents an overview of the corresponding criteria for evaluating and comparing alternative solutions to meet the stakeholder requirements. Section 4.2 introduces the development of the GIS based algorithm that integrates the requirement criteria in section 4.1 as well as the development of user-friendly interfaces. Both the system framework and key subroutines are illustrated. Section 4.3 presents the application of the algorithm to a real world case study, the city of Leicester, to demonstrate the practical applicability of the algorithm. A sensitivity analysis is presented in section 4.4 to demonstrate the impact of the different requirement criteria on the location and number of generated candidate solutions. Section 4.5 provides a summary of the findings.

4.1 Overview of the Rail Transit Station Location Requirements

As depicted in table 3.1, the intent of the identified requirements included in the planning framework can be translated into specific practical criteria for formulating the model and carrying out the evaluation. For each of the three interested parties, the number and type of criteria are designated based on a comprehensive literature review and the factors that influence the identified party's requirements (i.e., sub-objectives). While the consideration of the requirements helps to attain the stakeholders' main objectives, some of the identified requirement criteria, such as train travel speed and station density can only be integrated into the evaluation framework when the line network connecting the stations has been generated. Hence, they are not considered at this stage as this feasibility study addresses only the station sites regardless of how the stations contribute to the final configuration of the system. The requirements in Table 3.2 are used in this section to determine the feasibility of station locations and integrated into the proposed GIS based algorithm. The table shows that the community requirement criteria as more diverse than the passenger and operator requirements, although overlap significantly with them. There are two main reasons for this. Firstly, the community objectives, in addition to the passenger and operator perspectives, encompass all possible impacts of the planned system on the entire population of the served area and land use pattern. Secondly, long-term objectives such as improvement of economic productivity, environmental conditions and energy conservation are held only by the community; neither passengers nor operators are directly concerned with them. The GIS based algorithm is designed to accommodate all the requirement criteria depicted in Table 3.2 and eliminates the overlapping requirements among the three stakeholders. The following two sections detail how the algorithm integrates these requirement criteria, performs the search and computes the desired solutions.

4.2 The GIS Based Algorithm Framework

Figure 4.1 presents the system framework of the proposed GIS based algorithm that systematically integrates all the requirement elements illustrated in table 3.2 in order to measure the feasibility of station sites and generate a candidate pool of potential stations. The system consists of the following five principle modules:

1- Input Module: this is employed by rail transit planners to define: (1) station size, (2) passenger, operator, and community related parameters and (3) desired threshold values of the identified requirement criteria.

2- Station Generation Module: this divides the study area into grids based on the station size inputs. It is assumed that each of the generated grids represents a potential location of a rail station.
3- Station Location Evaluation Module: this module evaluates each potential station location generated from the Station Generation Module with respect to the identified requirement criteria presented in Table 3.2 using the input values of the passenger, operator and community related parameters (discussed in Section 3.3).
4- Station Location Designation Module: the aim of this module is to designate stations that satisfy the desired conditions set by the requirements of the three stakeholders after receiving the evaluated property of their corresponding locations from the Station Location Evaluation Module. The user defined threshold values of the requirement criteria from the input module represent the system desired conditions.
5- Output Module: this displays the attributed properties of the designated locations of the potential rail stations, and assists rail transit planners in examining the characteristics of the designated station locations.

All modules are integrated by exchanging data inside the ArcGIS environment. It is important to note that the algorithm uses walking distance as a base for measuring the requirement criteria. There are two main reasons for this. Firstly, the effects of rail stations on the various stakeholder requirements, such as people's attraction to using rail services, and improvement in economic and social activities are very strong within walking distance of stations, with little impact beyond the walking distance (Debrezion et al., 2007, Evans et al., 2007). Secondly, walking distance provides a standard basis for the comparative evaluation of the potential station locations. The following subsections elaborate in detail how each module incorporates the planning parameters, analyses them and then passes outputs to the other modules to, ultimately, screen the study area for feasible locations of stations and display the customized output of the candidate pool of potential stations.

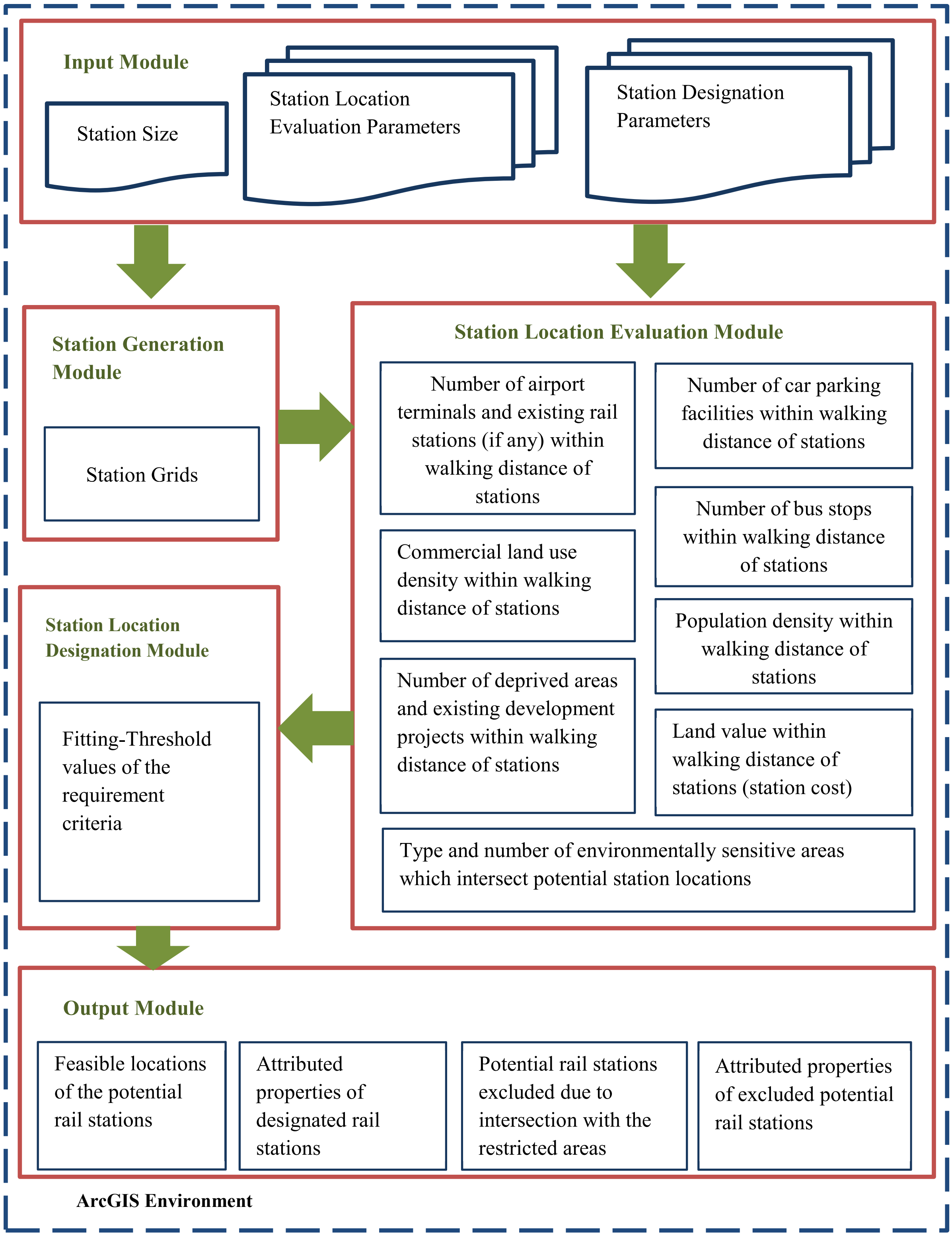


Figure 4.1: System Framework of the GIS-Based Algorithm

4.2.1 Input Module

This module consists of three interfaces for rail transit planners to input and adjust various planning parameters. These include station size, the parameters associated with the three interested stakeholders' requirement criteria and their corresponding desired threshold values. The parameters are defined in detail in chapter 3 and are discussed in this section.

4.2.1.1 Station Size

This interface enables rail transit planners to define the basic inputs for generating station grids which refer to the definition of station width and length and the geographical boundary of the area to be served, as illustrated in figure 4.2. The station size depends on the type of station to be planned (i.e., whether it is an underground or over-ground station), the number of rail lines passing through it and the size of the trains using it. However, since the number of lines passing through stations depends on the corresponding potential usage, and because this is very difficult to determine at this stage of the planning process, a fixed station size is used initially. Once the usage of the potential station has been estimated and the line network connecting the station has been generated, the size of that station is changed to be proportional to the number of lines passing through it, as discussed in section 5.4.3.4.

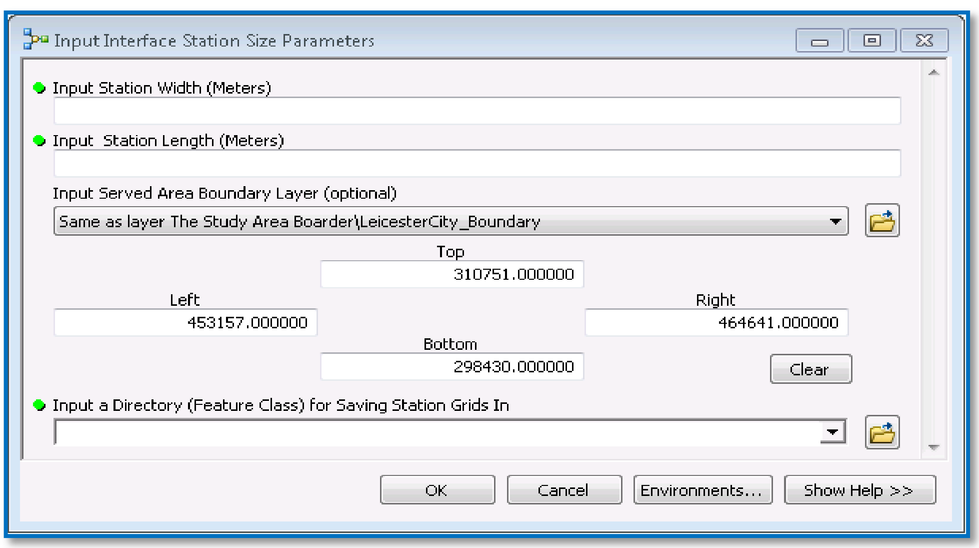


Figure 4.2: Input Interface: Station Size Parameters

4.2.1.2 Station Location Evaluation Parameters

This interface allows rail transit planners to define the basic inputs for evaluating locations of the generated station grids. The basic input settings for evaluating the potential locations of the generated station grids consist of the following parameters, as illustrated in figure 4.3:

1- Station availability - evaluation parameters: this corresponds to the value assigned to the walking distance to/from stations. As discussed in section 3.3.1.1, there is no accepted standard value for walking distance to rail stations for planning purposes. It varies with the type of area for which the system is planned (i.e. urban and suburban), the level and quality of service offered at rail stations and the built environment characteristics of the served areas. Based on relevant literature discussed in chapter 3, however, it can be concluded that a walking distance of between 400 and 800 metres to a station is the most effective range for satisfying the desired objectives.
2- Integration with other transport modes- evaluation parameters: these refer to the geographical locations of airport terminals, existing railway stations (if any), car parks and bus stops in the served area in the ArcGIS format and in separate feature classes.
3- Population coverage - evaluation parameters: these refer to the definition of the population density of each particular residential, commercial and business zone in the served area. This is also required to be defined in the ArcGIS format, feature class.
4- Passenger attraction - evaluation parameters: these are the geographical locations of the commercial land use areas in the served area. In this study, the commercial land uses include recreational areas, shopping malls, office complexes, industrial complexes, hospitals and university campuses. These areas need to be defined in the ArcGIS format, feature class.
5- Economic growth-evaluation parameters: these are the geographical locations of the economic growth potential sites which include deprived areas that need regeneration, vacant land with potential for future economic growth and existing development projects in the served area. As discussed in section 3.3.3.1, identification of these sites needs full coordination between transport and urban planning policies. Similar to the other evaluation parameters, these sites are required to be defined in the ArcGIS format and in separate feature classes.

6- Station construction cost - evaluation parameters: these involve the average land value in unit cost/unit area of each particular residential, commercial and business zone of the served area in the ArcGIS format, feature class.

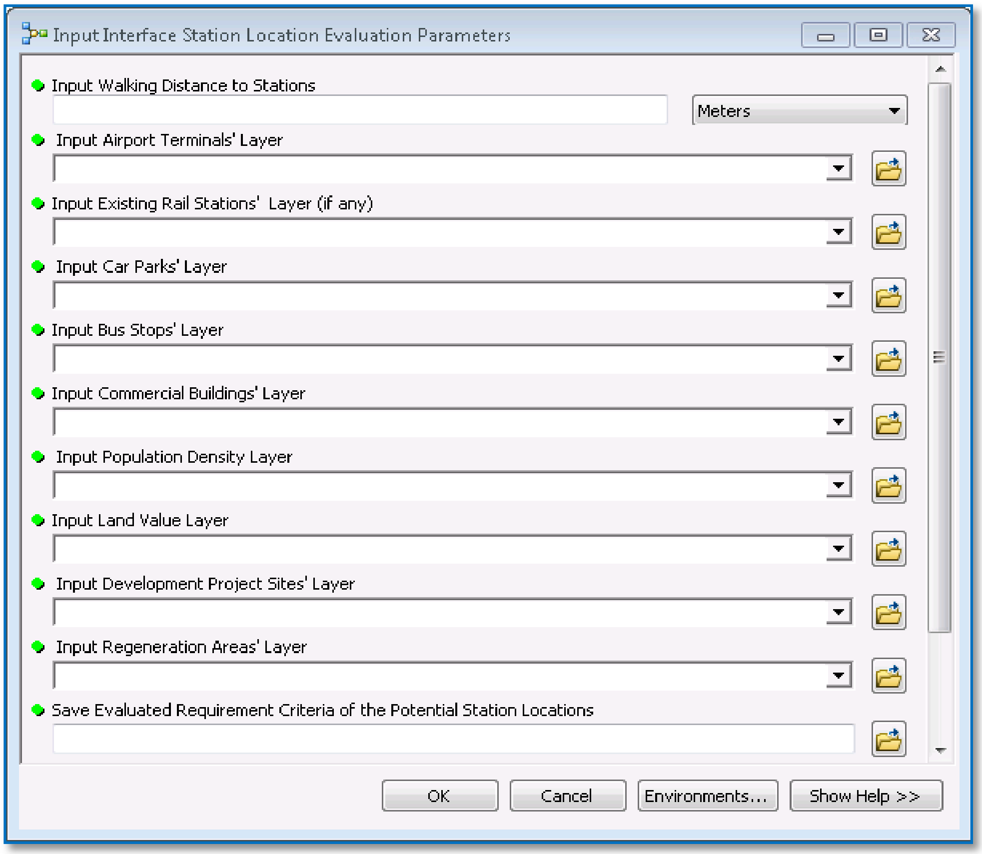


Figure 4.3: Input Interface: Station Location Evaluation Parameters

4.2.1.3 Station Designation Parameters

This interface is designed for rail transit planners to define the desired conditions for designating feasible locations of stations. These involve identification of environmentally sensitive areas that cannot be locations for stations, definition of the threshold values for each of the requirement criteria and setting of the minimum number of requirement criteria that must be satisfied by the generated station grid in order for it to be considered as a potential station. Figure 4.4 and 4.5 illustrate snapshots of this interface, which consists of the following three main input groups:

1- Environmentally sensitive areas: the identification of these areas depends largely on the type of rail system to be planned, and must account for all possible social and environmental issues that may arise due to the proposed rail system, as explained in section 3.3.3.2. This interface is designed to allow rail transit planners to input various sensitive area types, such as historic areas, sites of special scientific interest, lakes, rivers, forests and woodlands, as depicted in figure 4.4. Furthermore, this interface incorporates the flexibility to allow planners to consider the very specific environmental characteristics of the potentially served area. For example, countries involved in World War II have many undetonated bombs which still lie underground and leave a deadly legacy. These sites should be taken into account in planning a rail transit system, particularly when it is planned to be built underground.

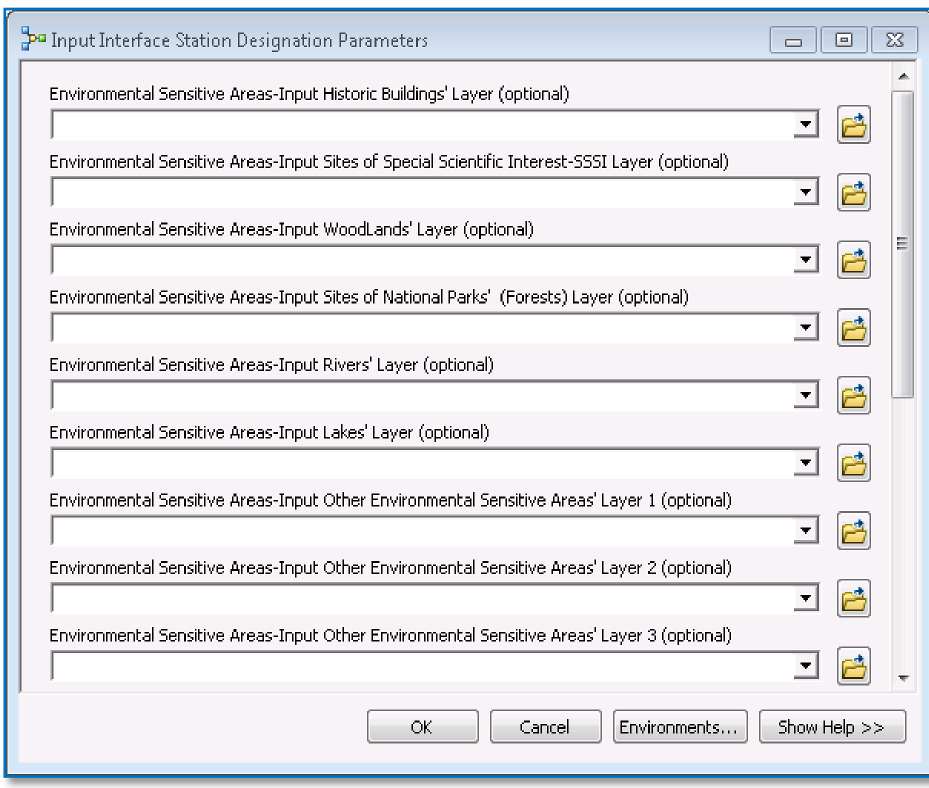


Figure 4.4: Input Interface: Environmental Sensitive Areas

2- Threshold values of the requirement criteria: these are critical for the evaluation of station location. These values are defined by rail transit planners depending on the demographic, topological, economic and land use pattern characteristics of the potentially served area, and include the following main inputs, as presented in figure 4.5.

i- Integration with other transport modes: threshold value to represent the required minimum numbers of airport terminals, existing railway stations, car parking facilities and bus stops within walking distance of stations.

ii- Population coverage: threshold value to represent the required minimum population density within walking distance of stations.

iii- Passenger attraction: threshold value to represent the required minimum commercial land use areas within walking distance of stations.

iv- Economic growth: threshold value to represent the required minimum deprived areas that need regeneration, vacant land areas with potential for future economic growth or existing development project areas within walking distance of stations.

v- Station construction cost: threshold value to represent the maximum average land value within walking distance of stations.

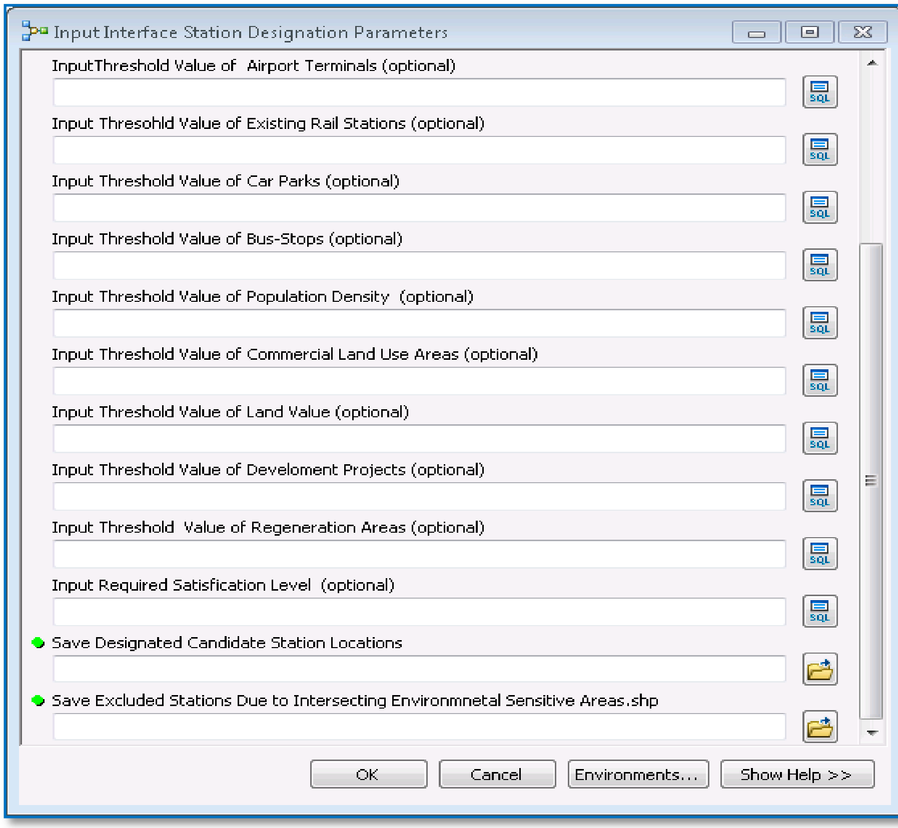


Figure 4.5: Input Interface: Requirement Criteria Threshold Values and Satisfaction Level

3- Number of satisfied requirement criteria: this involves the definition of a required minimum number of stakeholder requirement criteria that need to be satisfied by each of the generated station grids in order to be selected as a candidate potential station.

The decision regarding the threshold values is very critical since it may not only affect the number and locations of the generated candidate pool of potential stations, but also the entire configuration of the proposed rail system and subsequently the achievement of the desired system objectives. In general, the higher the threshold values, the better the satisfaction of the system requirements would be. On the other hand, setting high threshold values may result in significantly reducing the search space for the optimization algorithm, and consequently leading to a sub-optimal solution. The effects of the threshold values on the requirement criteria will therefore be assessed throughout the analysis process in order to investigate:

1- The potential risk associated with high threshold values for the station requirement criteria on the attainment of the entire system objectives due to excessively narrowing down the optimization searching space.
2- The impact of low threshold values on the performance of the optimization algorithm to generate an optimal system network that contributes to the acquisition of the desired system objectives

The decision regarding how many of the requirement criteria are to be considered in evaluating the potential served area for feasible station locations is of central importance in achieving the desired objectives of the major rail transit system planning stakeholders: passengers, operators and the community. Contradictions among some of the requirement criteria make it very difficult to find locations for stations that satisfy all requirements simultaneously. For instance, the usually low residential population size around business or commercial areas compared to residential areas complicates the joint consideration of criteria regarding population density and commercial land use areas. Setting a high threshold value for population density may result in exclusion of most of the commercial land use areas from the search space of feasible station locations due to low population density in these areas. In addition, this would affect the satisfaction of the economic growth requirements, specifically stimulation of economic activities in undeveloped areas, since population size in these areas is usually very low as well. Other conflicting criteria include land value and commercial land use areas. Obviously, land values associated with commercial areas are usually very high, therefore if areas with extremely high land values are avoided in order to reduce the construction cost of the system, most of the commercial areas may be excluded from the search space of potential station locations.

Attempting to meet all the requirements when scanning the potential served area for feasible station locations would therefore, be likely to lead to the generation of a very limited number of candidate stations. This may, in turn, not only result in excessive narrowing of the search space for the optimization algorithm, but may even result in an insufficient number of stations to generate a single rail line system. On the other hand, disregarding the contradicting requirement criteria would affect the success of the proposed system in fulfilling the desired objectives of the three interested stakeholders. Consequently, potential station locations must be determined as a trade-off between the various requirements of the three interested stakeholders by satisfying their corresponding requirements to a certain extent taking into account their relative importance regarding the acquisition of the primary objective for developing the proposed system.

4.2.2 Station Generation Module

The function of this module is to divide the potential served area into grids for use by the Station Evaluation and Designation modules. The grid size width and length values are defined by the rail transit planners, and should be equal to the actual size of rail stations since it is assumed that each generated grid represents a potential rail station. The number of the generated station grids is therefore, a function of the size of a potential served area and of the station size (width and length). The type of a rail system to be planned, number of tracks to pass through stations and train size constitutes key parameters in making up the size of a typical station structure. This module, however, can efficiently handle any station structure size, and this enhances its practical application for different planning conditions. The module consists of the **Generate Station Grid** function which is a subroutine to divide a served area into grids (Gi) based on the station size inputs, creating a layer (Ω_s) for stations. It is assumed that each of the generated grids represents a potential location of a rail station.

4.2.3 Station Location Evaluation Module

This module functions to evaluate the locations of the station grids generated from the Station Generation Module with respect to the requirement criteria for the three stakeholders. The evaluation results are then fed into the subsequent Station Designation Module so that they can be compared with the corresponding threshold values of the requirement criteria in order to identify a pool of feasible stations from the generated station grids. The Station Location Evaluation Module has the following key functions or subroutines:

- **Airport Terminal Frequency**: to count the number of airport terminals (if any) for each station grid within a defined search radius distance, i.e., walking distance to stations, and to assign the computed values to the corresponding station grid.
- **Rail Station Frequency:** to count the number of existing rail stations (if any) for each station grid within walking distance of stations, and to assign the computed values to the corresponding station grid.
- **Car Park Frequency:** to count the number of car parks for each station within walking distance of stations, and to assign the computed values to the corresponding station grid.
- **Bus Stop Frequency:** to count the number of bus stops for each station grid within walking distance of stations, and to assign the computed values to the corresponding station grid.
- **Population Density:** to compute the average population density for each station grid within walking distance of stations, and to assign the computed values to the corresponding station grid. The population density of each area/ city is usually estimated at ward or output area levels, with each ward/ output area denoting a particular area of the city. The population density is, therefore, not partitioned when a station grid's value is based on only one ward/output area of population density, i.e., when only one ward/output area lies within the defined walking distance of a station. In cases where more than one ward/ output area lies within walking distance of a station grid, a weighted average of population density is calculated based on the percentage contribution of each ward or output area to that walking distance radius.
- **Commercial Land Use Area:** to compute the commercial floor areas for each station grid within walking distance of stations, and to assign the computed value to the corresponding station grid. The commercial land use areas considered in this thesis include recreational areas, shopping malls, office complexes, industrial complexes, hospitals and university campuses.
- **High Economic Growth Site Area:** to compute the defined deprived areas that require revitalization, vacant land areas that have strong potential for economic growth and existing development project areas for each station grid within walking distance of stations, and to assign the computed values to the corresponding station grid.

- **Land Value:** to compute the average land value, in unit cost per unit area, for each station grid within walking distance of stations, and to assign the computed values to the corresponding station grid. Since land value is estimated at postcode level, when only one postcode block lies within a defined walking distance area of a station, the land value for that area is not partitioned. In cases where more than one postcode block lies within the walking distance area of a station grid, a weighted average is calculated based on the percentage contribution of each postcode block to that walking distance area.

4.2.4 Station Location Designation Module

This module is the core component of the proposed system framework. It implements comparative analysis and generates the candidate pool of potential stations that satisfy the desired conditions defined in the Input Module. The generated solution is then used in the optimization algorithm as the input for generating the final system configuration. This module consists of the following key functions:

- **Environmentally Sensitive Area Exclusion:** to find all station grids that intersect with environmentally sensitive areas, defined in the Input Module, and to exclude them from the search space for potential station locations. This is achieved by overlaying the generated station grid layer (Ω_s) and the environmental sensitive areas layers (Ω_i) to generate the feasible grid layer for stations(Ω_f).
- **Station Location Assessment:** to find all grids within the feasible station layer (Ω_f) that satisfy the conditions set by the threshold values of the requirement criteria and to weight them with binary values, as explained in chapter 3, through equations (3.2) to (3.9). The attributed weight to each condition represents the degree of importance attached to the corresponding requirement criteria. The relative importance of the requirements differs between different cities or served areas depending on the primary objectives of the intended system. In cases where, for instance, the major objective of building the system is for the development and regeneration of economic activities, the community requirements, specifically the criteria of economic growth, are weighted as more important than the other requirement criteria. In the

algorithm proposed here, however, equal weight is given to all the requirement criteria. Since a rail transit system is a major, high investment, permanent structure, it should not be solely linked to a single objective but also viewed in the broad context of the mobility, socio-economic, public transport and environmental improvements that it may offer. Note that the algorithm is designed to accommodate different weights for different requirement criteria in cases where the prioritization of a particular objective is necessary.

- **Station Location Designation:** to aggregate the weight of the requirement criteria that are satisfied by each grid and then to designate all grids that meet the satisfaction level defined in the Input Module. Following the procedure defined in the Station Location Assessment subroutine the attached weight of each requirement criteria is assumed to be equal to 1, the maximum weight that each station grid can gain is therefore 8. This is because the desired objectives of passenger, operator and community are broken down into eight requirement criteria (excluding the environmentally sensitive area requirement), as discussed in section 4.1. These aggregate values also reflect the total number of the requirement criteria satisfied by station grids. This is because when a station grid does not meet the desired condition of a particular requirement criterion, it will have a zero weight. For example, when the calculated aggregate weight of a station grid is 3 it means that particular grid meets the desired condition of 3 requirement criteria. The reason for aggregating the weights of the requirement criteria into a single value rather than into three different values, one for each particular stakeholder, is due to the large overlap between the three interested stakeholders' requirement criteria. If the aggregate weight of each of the stakeholder's requirement criteria were computed individually, the overlapped requirement criteria would be counted more than once.

4.2.5 Output Module

Once all the station grids have been assigned with binary values according to the procedure defined in section 4.2.4, the candidate pool of potential station locations and associated properties is generated and displayed in the ArcGIS environment. In addition, the Output Module generates the property table of the station grids that were excluded from the candidate potential station locations due purely to environmental and social considerations. This helps the rail transit planner to assess the effect of these areas on the number and locations of the generated potential solutions. Figures 4.6 to 4.8 show the snapshots of the Output Module display.

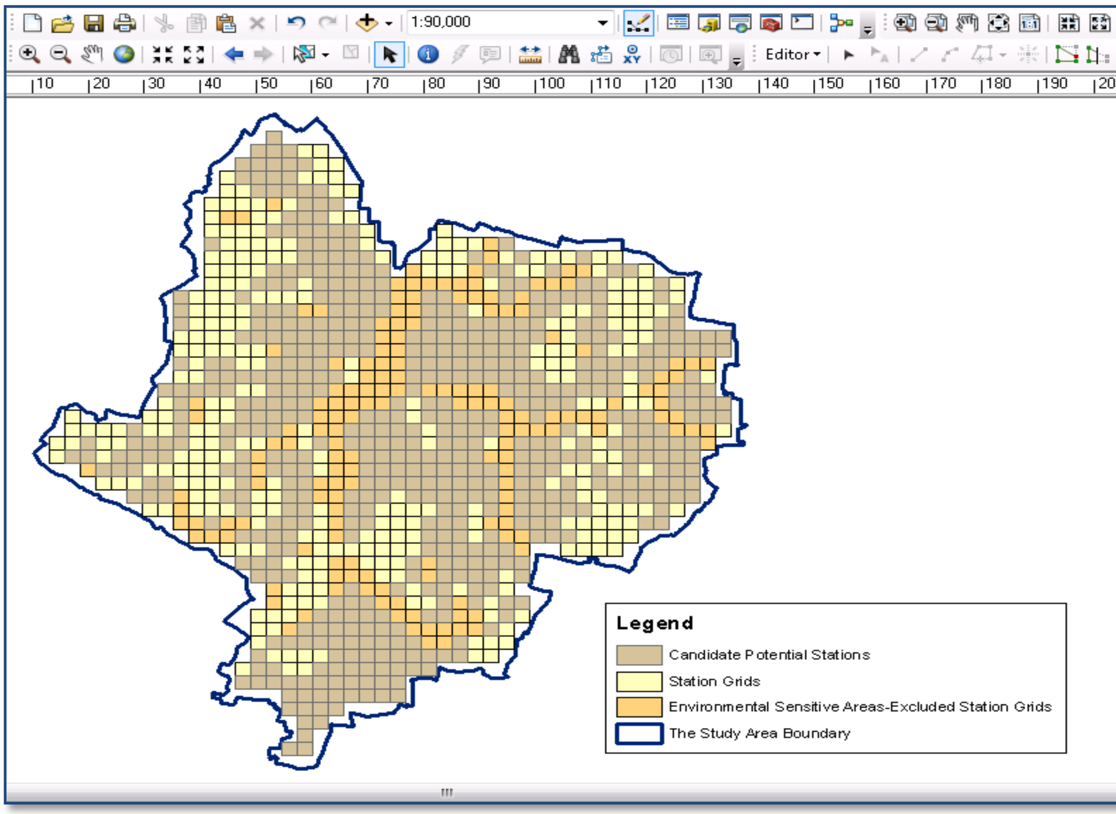
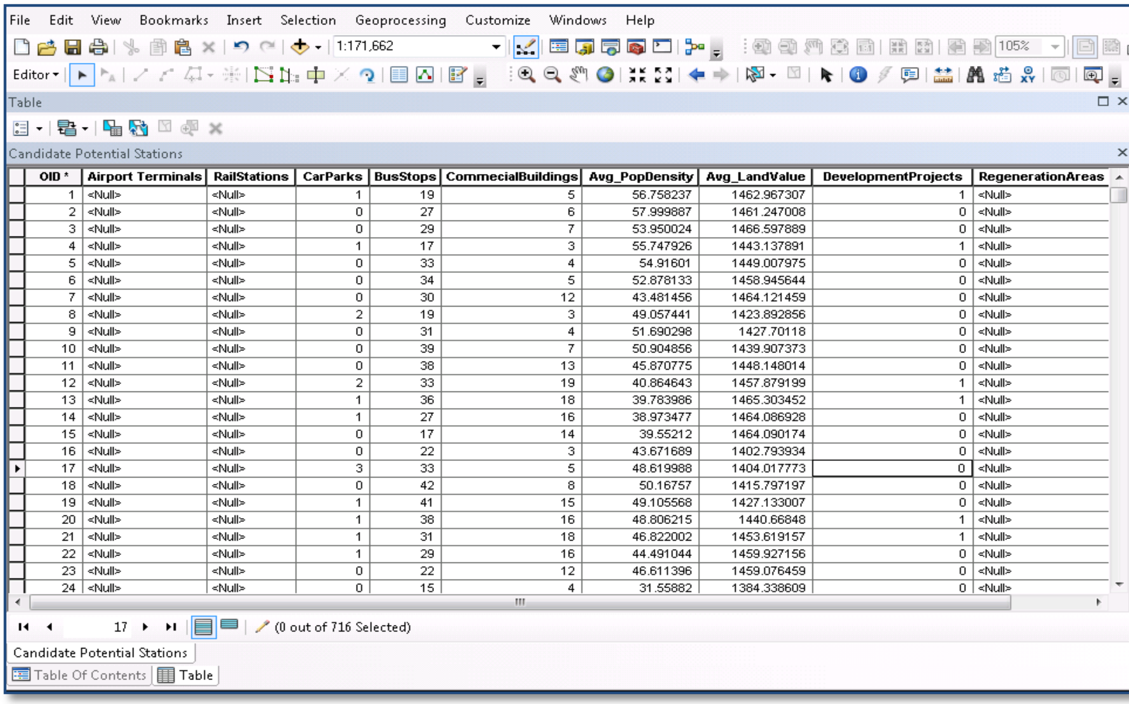


Figure 4.6: Candidate Pool of Potential Station Locations Output

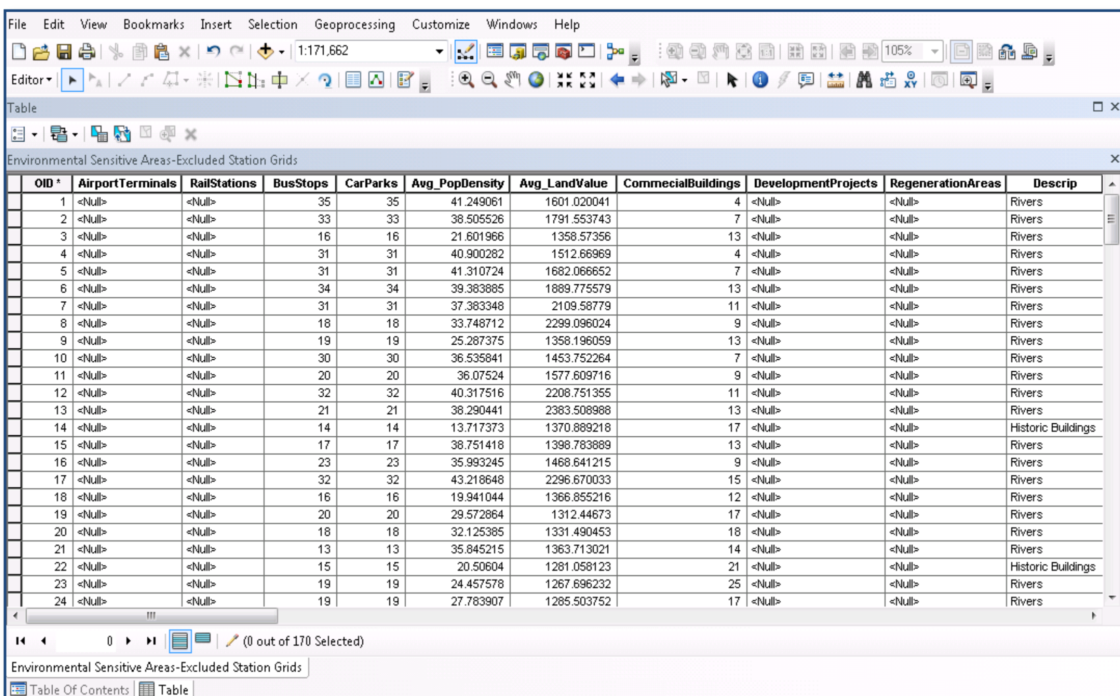


Candidate Potential Stations

OID *	Airport Terminals	RailStations	CarParks	BusStops	CommecialBuildings	Avg_PopDensity	Avg_LandValue	DevelopmentProjects	RegenerationAreas
1	<Null>	<Null>	1	19	5	56.758237	1462.967307	1	<Null>
2	<Null>	<Null>	0	27	6	57.999887	1461.247008	0	<Null>
3	<Null>	<Null>	0	29	7	53.950024	1466.597889	0	<Null>
4	<Null>	<Null>	1	17	3	55.747926	1443.137891	1	<Null>
5	<Null>	<Null>	0	33	4	54.91601	1449.007975	0	<Null>
6	<Null>	<Null>	0	34	5	52.878133	1458.945644	0	<Null>
7	<Null>	<Null>	0	30	12	43.481456	1464.121459	0	<Null>
8	<Null>	<Null>	2	19	3	49.057441	1423.892856	0	<Null>
9	<Null>	<Null>	0	31	4	51.690298	1427.70118	0	<Null>
10	<Null>	<Null>	0	39	7	50.904856	1439.907373	0	<Null>
11	<Null>	<Null>	0	38	13	45.870775	1448.148014	0	<Null>
12	<Null>	<Null>	2	33	19	40.864643	1457.879199	1	<Null>
13	<Null>	<Null>	1	36	18	39.783986	1465.303452	1	<Null>
14	<Null>	<Null>	1	27	16	38.973477	1464.086928	0	<Null>
15	<Null>	<Null>	0	17	14	39.55212	1464.090174	0	<Null>
16	<Null>	<Null>	0	22	3	43.671689	1402.793934	0	<Null>
17	<Null>	<Null>	3	33	5	48.619988	1404.017773	0	<Null>
18	<Null>	<Null>	0	42	8	50.16757	1415.797197	0	<Null>
19	<Null>	<Null>	1	41	15	49.105568	1427.133007	0	<Null>
20	<Null>	<Null>	1	38	16	48.806215	1440.66848	1	<Null>
21	<Null>	<Null>	1	31	18	46.822002	1453.619157	1	<Null>
22	<Null>	<Null>	1	29	16	44.491044	1459.927156	0	<Null>
23	<Null>	<Null>	0	22	12	46.611396	1459.076459	0	<Null>
24	<Null>	<Null>	0	15	4	31.55882	1384.338609	0	<Null>

(0 out of 716 Selected)

Figure 4.7: Attribute Table of the Candidate Potential Station Locations



Environmental Sensitive Areas-Excluded Station Grids

OID *	AirportTerminals	RailStations	BusStops	CarParks	Avg_PopDensity	Avg_LandValue	CommecialBuildings	DevelopmentProjects	RegenerationAreas	Descrip
1	<Null>	<Null>	35	35	41.249061	1601.020041	4	<Null>	<Null>	Rivers
2	<Null>	<Null>	33	33	38.505526	1791.553743	7	<Null>	<Null>	Rivers
3	<Null>	<Null>	16	16	21.601966	1358.57356	13	<Null>	<Null>	Rivers
4	<Null>	<Null>	31	31	40.900282	1512.66969	4	<Null>	<Null>	Rivers
5	<Null>	<Null>	31	31	41.310724	1682.066652	7	<Null>	<Null>	Rivers
6	<Null>	<Null>	34	34	39.383885	1889.775579	13	<Null>	<Null>	Rivers
7	<Null>	<Null>	31	31	37.383348	2109.58779	11	<Null>	<Null>	Rivers
8	<Null>	<Null>	18	18	33.748712	2299.096024	9	<Null>	<Null>	Rivers
9	<Null>	<Null>	19	19	25.287375	1358.196059	13	<Null>	<Null>	Rivers
10	<Null>	<Null>	30	30	36.535841	1453.752264	7	<Null>	<Null>	Rivers
11	<Null>	<Null>	20	20	36.07524	1577.609716	9	<Null>	<Null>	Rivers
12	<Null>	<Null>	32	32	40.317516	2208.751355	11	<Null>	<Null>	Rivers
13	<Null>	<Null>	21	21	38.290441	2383.508988	13	<Null>	<Null>	Rivers
14	<Null>	<Null>	14	14	13.717373	1370.889218	17	<Null>	<Null>	Historic Buildings
15	<Null>	<Null>	17	17	38.751418	1398.783889	13	<Null>	<Null>	Rivers
16	<Null>	<Null>	23	23	35.993245	1468.641215	9	<Null>	<Null>	Rivers
17	<Null>	<Null>	32	32	43.218648	2296.670033	15	<Null>	<Null>	Rivers
18	<Null>	<Null>	16	16	19.941044	1366.855216	12	<Null>	<Null>	Rivers
19	<Null>	<Null>	20	20	29.572864	1312.44673	17	<Null>	<Null>	Rivers
20	<Null>	<Null>	18	18	32.125385	1331.490453	18	<Null>	<Null>	Rivers
21	<Null>	<Null>	13	13	35.845215	1363.713021	14	<Null>	<Null>	Rivers
22	<Null>	<Null>	15	15	20.50604	1281.058123	21	<Null>	<Null>	Historic Buildings
23	<Null>	<Null>	19	19	24.457578	1267.696232	25	<Null>	<Null>	Rivers
24	<Null>	<Null>	19	19	27.783907	1285.503752	17	<Null>	<Null>	Rivers

(0 out of 170 Selected)

Figure 4.8: Attribute Table of the Excluded Station Grids Due to Intersecting Environmental Sensitive Areas

4.3 Case Study

This section examines the effectiveness of the proposed model, specifically its GIS based algorithm part, by applying it to a real world case study. Although the model is designed in such a way that it can be used for any city in the world, four main criteria are used to select the case study in this thesis: (1) population size of the city, (2) special structure or form of the city, (3) availability of a rail transit system in the city, and (4) geographical features of the city. These factors are very important in order to determine the volume of travel along main corridors, or throughout the city, and to determine whether the investment involved in establishing a rail transit system is justified or not. It is widely accepted that a fairly large population size is an essential requirement for the success of a rail transit system. However, as Vuchic (2007) points out, the construction of a rail transit system is largely influenced by various local factors in relation to topographical, political and financial aspects, and therefore there is no standard threshold value of the ideal population size that justifies the construction of such a system. Very successful rail transit systems have been built in many cities having different population sizes. For example, the population size of Stockholm was 750,000 when its rapid rail transit system (metro) began operation, while the populations in Lisbon and Rotterdam were 900,000 and 700,000 respectively, and Nurnberg and Oslo built their systems before their population reached 500,000. Also, it is important to note that the threshold for use of the other rail transit system types, especially light rail transit, is considerably lower than that for rapid rail transit. Light rail transit systems are usually planned for cities with 150,000 to 300,000 populations. Many cities with population ranges between 500,000 and 2 million, and some even significantly larger, utilize both light and rapid rail transit systems, examples include London, Amsterdam, Milan, San Francisco and Toronto (Vuchic, 2007). Based on these observations it can be concluded that a city with a population greater than 300,000 would have the potential to support the construction of a rail transit system. Accordingly, many cities in the UK, specifically in England, could be candidates for a case study, such as London, Manchester, Nottingham, Leeds, Leicester, Newcastle, Sheffield and Birmingham.

The spatial structure or form of the city, which refers to the extent to which commercial and business activities are concentrated in a single dominant centre (i.e., a monocentric city) or in multiple centres (i.e., a polycentric city), significantly contributes to the success of building a rail transit system. In general, most cities in the world have a mixed structure, a proportion of trips are radial and follow the monocentric structure while others have random origins and destinations and follow the polycentric structure. That is, most cities are monocentric or polycentric by degree only, although this includes some that are dominantly monocentric and others that are dominantly polycentric (Bertaud, 2002). International experience from existing rail transit based systems shows that such systems are generally very successful when implemented in dominantly monocentric cities, due to the pattern of the trips (Cervero and Guerra, 2011, Bertaud, 2002). Tokyo, London, New York and Paris are good examples of such cities, which to some extent explain why their rail transit systems are successful. Each of these cities has central business districts with more than 750,000 jobs which significantly contribute to the success of the rail transit systems in utilizing their maximum capacities and increasing public transit mode shares (Mohan, 2005). Accordingly, this criterion argues for the exclusion of some UK cities with weekly monocentric structures (e.g. Leeds and Sheffield) from the pool of candidate cities (The Northern Way, 2009).

The third criterion (i.e., the current availability of a rail transit system) is used to further refine the pool of the candidate cities, with only those cities that currently do not have any rail-based system supplementing their existing public transport modes being retained in the pool. Although the proposed model can in principle be used to extend existing rail systems, for the purposes of the case study the plan of a new multi-line rail system is selected to better demonstrate the capabilities of the model.

The fourth criterion, which is the geographical features of the city, is very important to determine whether the investment in establishing a rail transit system is justified or not. Complex geographical features, such as hilly terrain, narrow valleys, and bodies of water make the construction of transportation facilities very costly, encouraging the construction of a high capacity transport mode, such as a rapid rail transit system, as a logical choice (Vuchic, 2007).

It is worth noting that the last two criteria (i.e, the lack of an urban rail system and the presence of complex geographical features) are both useful for confirming the ability of the proposed model to provide a comprehensive consideration of the various requirements of areas characterized by complex geographical features, and in generating potential solutions for particular corridors and the entire city.

Among the pool of potential case studies, the city of Leicester has been selected as it satisfies the above four criteria, especially the last two criteria which are important in demonstrating the model capabilities. The City of Leicester is located in the East Midlands of England, and is the county town of Leicestershire, as shown in figure 4.9. The size of the city area is 73.32 km^2 and its population of approximately 330,000 is the highest in the East Midlands region. The city has vibrant commercial, recreational, cultural and educational centres, making it a lively and attractive place. It is also directly linked with the rest of the country by strategic road and rail networks through its position at the heart of the country. Buses, however, are the only public transport mode that services the city itself. Building a rail transit network would, therefore, be very likely to enhance mobility, environmental protection and economic activity. The case study is defined by the city's geographical features, such as land use patterns, existing transport networks, and property and population distribution. The preparation of the required input data for the algorithm and the generated potential solution results are presented in the next subsections. It should be noted that two assumptions are made while applying the algorithm. Firstly, it is assumed that the type of the proposed rail system is underground/metro and; secondly, that the stations are built above ground while the line network linking the stations is built in deep tunnels. The proposed integrated optimization model can be applied to other types of rail transit system planning, (e.g., over ground rail systems) with some modifications (explained in section 5.4.3) to the construction cost parameters presented in section 3.4.2.1.

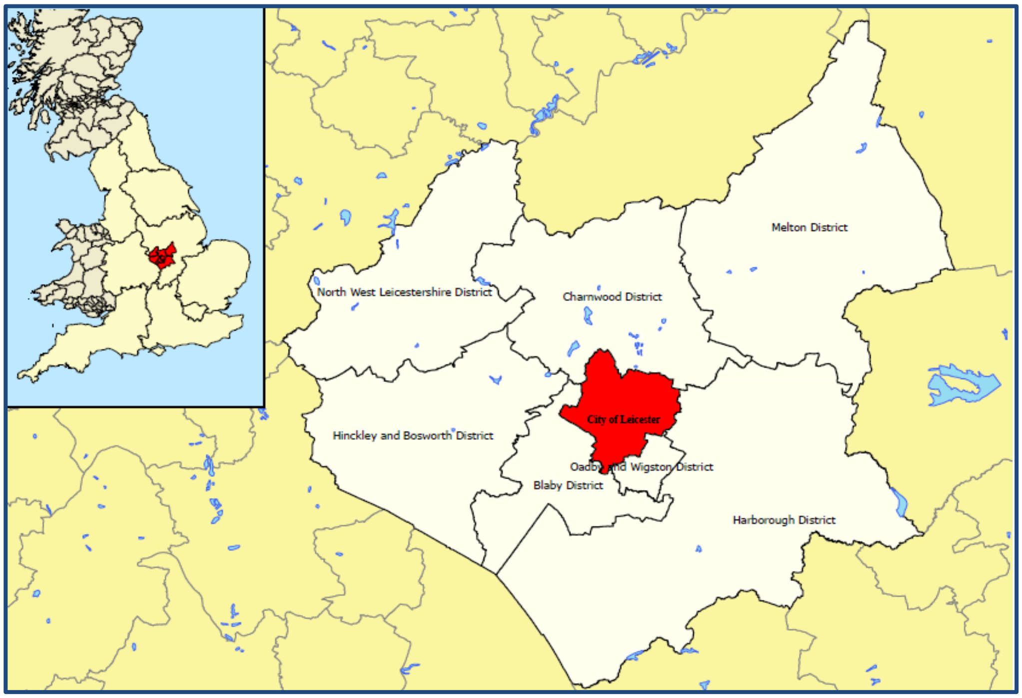


Figure 4.9: Illustration of the Case Study Location –City of Leicester

4.3.1 Data Pre-processing

The required datasets for the study area were obtained from several agencies and incorporated into the GIS background database. These datasets include the City of Leicester boundary, land use pattern, existing transport network, census data, land value and future economic action plan. The required input data, described in section 4.2.1, were extracted from these datasets and retrieved from the ArcGIS database in order to apply the proposed algorithm.

4.3.1.1 Land Use Pattern

The land use pattern data for the City of Leicester were obtained from the Ordnance Survey of Great Britain in the ArcGIS format, feature class, which contained geographic features of the city boundary, water areas, woodlands, forests, and recreational, residential, commercial, and institutional areas. These datasets were then restructured into the following group feature classes and layers, in order to match the structure of the proposed algorithm's required inputs, as defined in section 4.2.1.

1- Ceremonial boundary: this group contains only the City of Leicester boundary layer with the respective attributes which include features' identifier code, geometric shape and geometric area in square metres.

2- Commercial land use areas: this group contains the commercial land use layer with the associated attributes including identifier code, geometric shape and description of each feature. The features in this layer consist of shopping malls, museums, government offices, historic buildings, parks, sports fields, university campuses, schools and hospitals. However, this dataset, is made up of point features and therefore, their corresponding areas could not be calculated. Therefore, instead of using commercial land use area in the evaluation process as originally intended (see equation 3.7), their respective frequencies were used in the evaluation process, i.e. the total number of commercial buildings within walking distance of stations were computed and used in equation (3.7).

3- Environmental sensitive areas: this group contains layers of all the land use types that are required to be excluded from the search space for station locations due to environmental and social concerns. It consists of individual layers for historic buildings, woodlands, national parks and areas which involve rivers, lakes and canals that are collectively built up to provide land use pattern data. However, the woodland and national park layers are empty due to the lack of these land use types in the city. In addition, since the proposed system is planned as to be underground, a layer for unexploded bomb sites is added to this group. Although every effort was made to obtain information from the relevant government entities on the positions of unexploded bombs that had fallen on the city during World War II, this information could not be obtained. As a result, hypothetical positions of unexploded bombs were plotted on the city map which can be replaced with the real locations when obtained. Creating such hypothetical data was only to examine the capability of the proposed algorithm in accommodating a wide range of environmental and social requirements.

4.3.1.2 Existing Transport Network

Data on existing transport facilities, including airport terminals, railway lines and stations, car parks and bus stops were obtained from the Ordnance Survey of Great Britain and Leicester City Council. These data were restructured into the following group layers based on the structure of the proposed input module, defined in section 4.2.1.

1- Airport terminals: this group contains a layer for airport terminals with the associated attributes which provide the general geometric shapes and a description of the involved features.
2- Railway system: this group comprises railway line and station layers with the corresponding attribute tables which provide general geometric shapes and a description of the involved features. Rail routes run north-south through the City of Leicester, along the route known as the Midland Main Line, connecting the north-south cross country routes, and they serve Leicester Central railway station which lies on the eastern side of the city centre.
3- Car parks: this group contains a layer for car parks with the associated attributes which include identifier code, geometric shape and description of each feature. However, information about the car parks in the entire city could not be obtained. Therefore, this dataset does not cover all car parking locations in the city.
4- Bus stops: this group contains a layer for bus stops and associated attributes which include identifier code, geometric shape and a description of each bus stop.

4.3.1.3 Census Block Data

Census data were obtained from the United Kingdom Data Service (UK Census, 2011). Year 2011 census data for the City of Leicester were used, which included population density. The population density was calculated by dividing the number of usual residents to Output Area (OA). The output areas (OAs) are the base unit for census data releases, which were introduced in UK by the Office for National Statistics in the 2001 Census. They were built from clusters of adjacent unit postcodes that have similar population sizes and homogenous tenure of households and dwelling types using the data from the 2001 Census. The total numbers of output areas created for the City of Leicester were 890 blocks, and the population density was computed for each block.

4.3.1.4 Land Values

This dataset was obtained from the Land Registry for England and Wales, and consists of average property prices at postcode sector level in Q3 2014. The Land Registry (LR) publishes on a monthly basis, data on price paid for residential properties. It should be noted the data available was for Q3 2014. The price paid data includes properties that were sold at the full market value and lodged with LR for registration since 1995. The dataset excludes non-residential properties. This dataset, therefore, does not explicitly capture the required average land price data. Due to lack of data about the values of non-residential properties such as commercial and recreational, it is assumed that the LR data represents the average land values. Although the LR dataset mainly captures residential properties, it is considered that it reflects the land price pattern of the study area which should be sufficient for examining the effectiveness of the proposed algorithm.

4.3.1.5 Future Economic Action Plan

The strategic action plan for developing and unlocking economic growth in Leicester City was obtained from Leicester City Council. This economic plan outlines a substantial investment plan by the city council and its partners to create an environment for developing and vitalizing the city's economic growth between the period 2012 and 2020. The major theme of this plan is the establishment of four business investment areas in parts of the city with strong potential for growth. Each of these investment areas will focus on a specific business sector: food and drink manufacturing in the north east area of the city; innovation and technology relating to space and environmental technologies in the north area of the city centre ; creative industries centred on the "St George's Cultural Quarter" within the south east edge of the city centre. The fourth business investment area is the city centre itself, with the aim of supporting existing retail and encouraging the development of new leisure facilities and office space for professional and business services. In order to incorporate the need to accelerate economic growth into the decision-making platform, a GIS database for these four sector-specific business investment areas was created with respect to the corresponding relative geographical locations and proposed workspaces presented in the Leicester City Council plan.

4.3.2 The Algorithm Output Results

After the preparation of the required datasets in section 4.3.1, and considering an underground rail system, the following input parameters were applied to generate the candidate pool of potential rail stations:

1- **Station size:** the size of a typical, single-level track, underground/metro station structure is 16x25x300 metres for the respective depth, width and length (Kaul, 2010). It is assumed, that the station concourse hall will be built at ground level to give the system a more attractive view. Therefore, a station size of 100x100 meters was adopted in the Input Module. These dimensions represent the station concourse hall width and length. Although it is larger than the required typical dimension for the station concourse areas, it is considered that accommodating some retail and coffee shops within this area would have many positive effects, such as increased system attractiveness and passenger convenience. In addition, more than one line may pass through some of the candidate stations after the line network is generated. This obviously requires a larger dimension for the station structure so that it can accommodate additional services and associated facilities. Examples of such stations include Victoria and Kings Cross rail stations in London

2- Based on this station size, the city has a total of 7027 grids which represent potential rail station locations, and for each of these grids the defined requirement criteria were measured within a 500 metre walking distance radius from the stations. The choice of this value was based on the literature review (section 3.3.1.1).

3- **Threshold values:** the 60th percentile of each of the calculated requirement criteria values was defined as its respective threshold value. For example, the threshold value of the bus stop requirement criteria was found by first listing all the generated station grids in ascending order according to the number of bus stops located within their walking distance area. The 60th percentile of this list was then identified (i.e. 0.6 x 7027) and the number of bus stops in the particular station grid falling at this percentile in the ascending list was set as the threshold value for bus stops for the model.

4- **Satisfaction level:** The satisfaction level was set to be 50% or greater, which implies that the generated potential solutions must satisfy the threshold values of at least half of the station location requirements criteria presented in figure 4.1.

The choice of the threshold values and satisfaction levels was related to the potential number of candidate stations. Due to contradictions in some of the requirement criteria, adopting very high values for these parameters would cause excessive narrowing of solution alternatives that might result in losing good solutions. A sensitivity analysis was, therefore, carried out in section 4.4 to determine the degree of impact the different threshold values and satisfaction levels on the size of the generated solutions.

After defining the proposed conditions for screening the study area and generating the candidate pool of potential station locations, 3690 feasible station locations were found out of 7027 station grids, as shown in figure 4.10. The total number of infeasible station locations due to intersection with the environmentally protected areas (i.e., restricted zones) was 683, 367 of which otherwise satisfied the defined desired conditions for potential station locations.

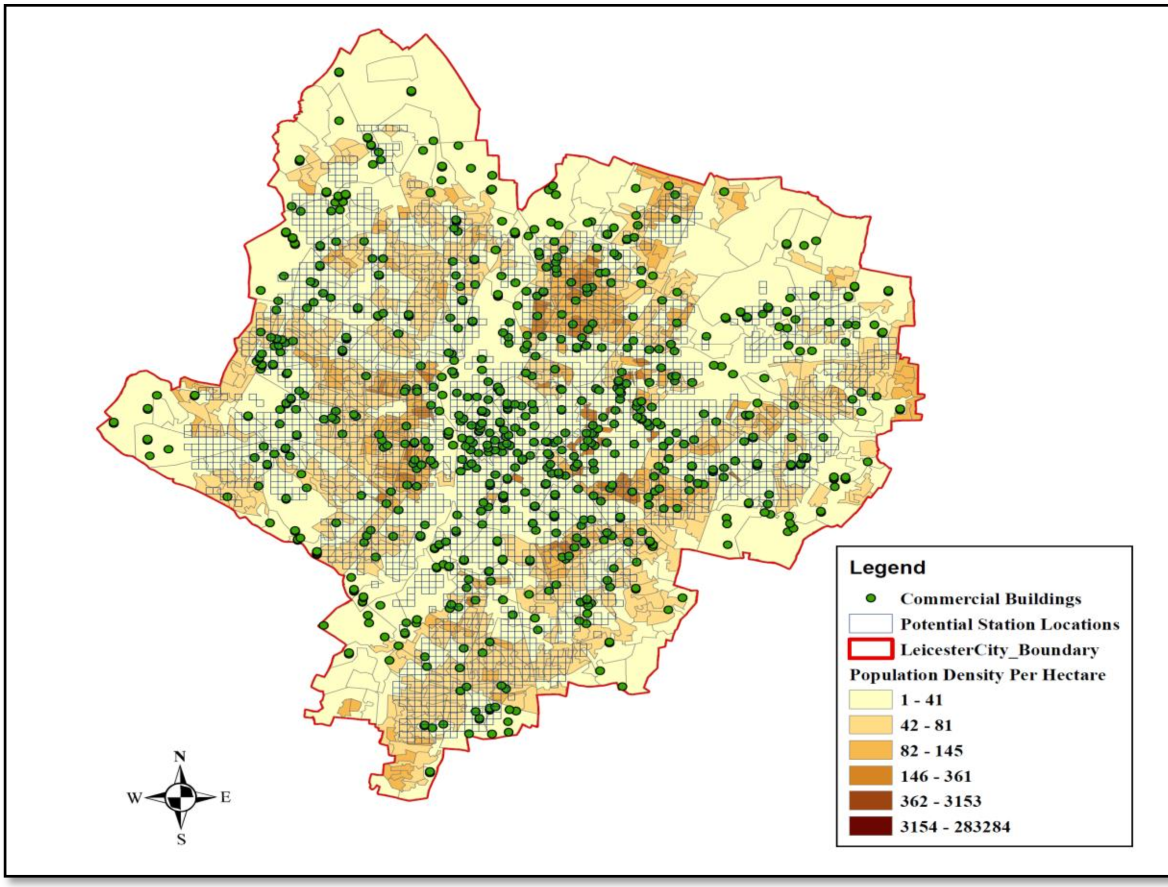


Figure 4.10: Generated Feasible Station Location Solutions

The results shown in figure 4.10 reveal that the candidate stations mainly cover the most densely populated and potential economic growth zones of the city. In addition, areas with very high land values and environmentally sensitive areas were excluded from the search space. For simplicity, some features such as sites of economic growth, bus stops, land values and restricted zones are not presented in the figure. This candidate pool of feasible station locations are used in chapter 5 as a building block for generating a network of lines linking the candidate stations to obtain the final architecture of the proposed rail system.

4.4 Sensitivity Analysis

This section investigates the effect of the threshold values and satisfaction levels of the requirement criteria for the three stakeholders' objectives on the number of generated feasible station location solutions. This is achieved by conducting two categories of sensitivity analysis. The first category examines the effect of threshold values, while the second category investigates the effect of the satisfaction levels on the number of feasible station locations.

4.4.1 Effect of the Requirement Criteria Threshold Values

The sensitivity analysis on the effect of threshold values compares nine different threshold values defined by different percentiles of the various requirement criteria values of the three interested stakeholders, as illustrated in figure 4.11.

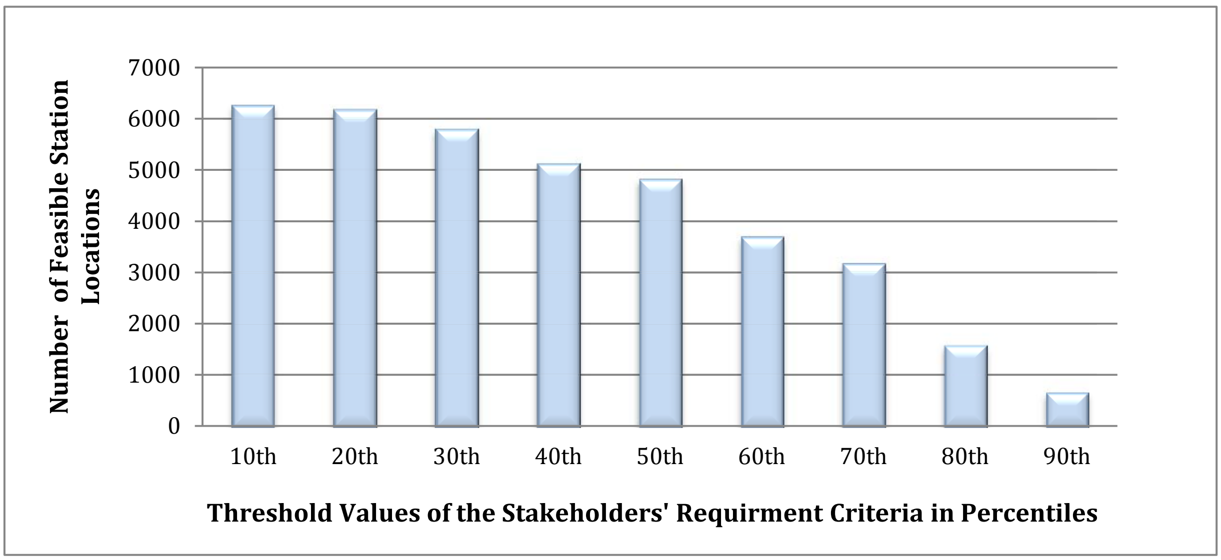


Figure 4.11: Number of Feasible Station Locations at Different Threshold Values

The comparison results show that increasing the threshold values of the requirement criteria clearly reduces the number of generated stations. This reduction becomes more significant at threshold values greater than the 50th percentile. The number of the generated solutions, for example, dropped to 4803 from 6241 (about 24%) when the threshold values of the requirement criteria increased from the 10th to the 50th percentiles. However, when the threshold values increased from the 50th to the 90th percentiles, the number of feasible stations reduced to 637 from 4803 (about 87%). There are two possible reasons for this, the first due to the spatial distribution of population and land use in the study area/city. The more uniform the distribution of the population and land use structure, the less the sensitivity of high threshold values to the number of generated solutions.

The second reason is the conflict between some of the requirement criteria, such as between commercial land use areas and either population density or land value. The effect of the conflicting requirements on the number of generated solutions increases considerably with their respective threshold values. For example, since the population density in areas with high levels of commercial land use is significantly less than that in residential areas, increasing the thresholds of the commercial land use areas and the population density requirements simultaneously, makes the satisfaction of both difficult resulting in a considerable decrease in the number of generated solutions. Similarly, due to the correlation of high land values with business areas, setting higher thresholds for the commercial land use areas and for land values significantly decreases the number of generated solutions.

4.4.2 Impact of the Requirement Criteria Satisfaction Levels

This section investigates the effect of the satisfaction levels of the requirement criteria on the number of feasible station locations. Nine threshold values (defined by different percentiles) were examined at four different satisfaction levels: 25%, 50%, 75% and 100%. The analysis results show that the number of generated solutions decreased significantly as the required satisfaction levels were increased, as illustrated in figure 4.12. However, this decrease, becomes very significant beyond the 25% satisfaction level. At the 25% satisfaction level, the number of generated solutions starts to reduce only beyond the 80th percentile threshold values, whereby the number of potential station locations are reduced to 4917 from 6344 (about 22.5%) after raising the threshold values from the 10th to 80th percentile.

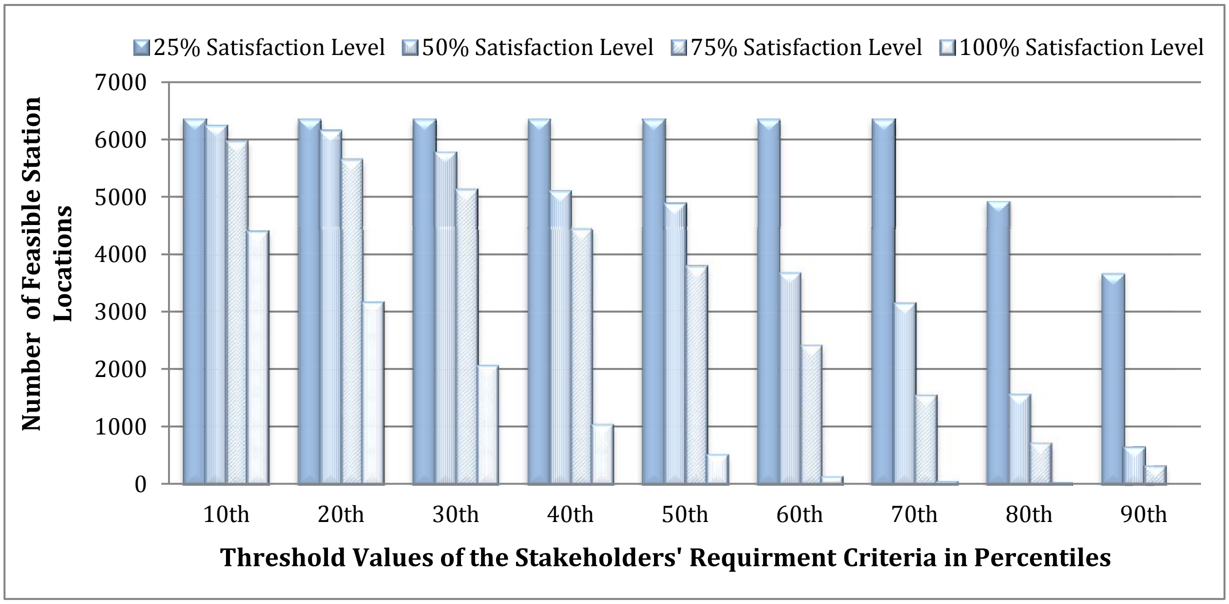


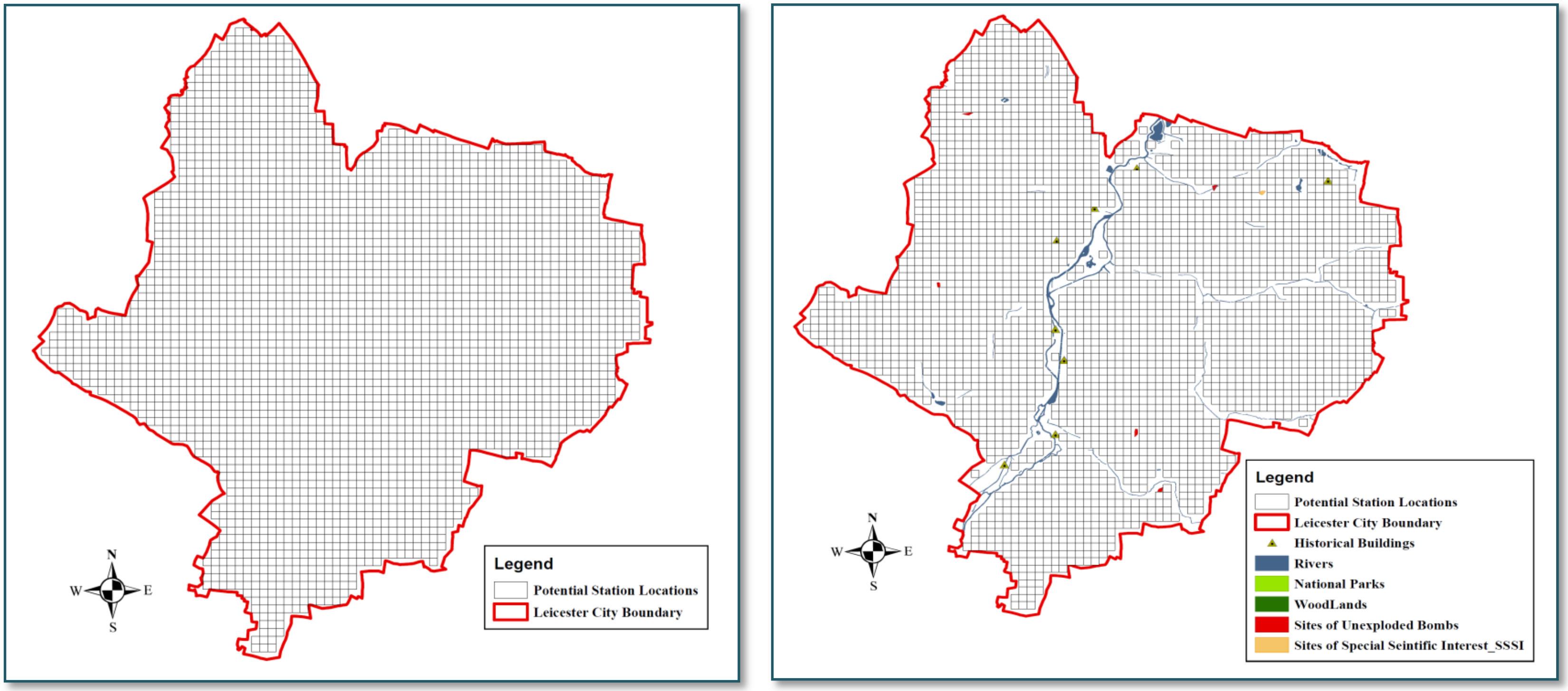
Figure 4.12: Number of Feasible Station Locations at Different Satisfaction Levels

Beyond this satisfaction level, however, the number of generated solutions decreased rapidly when the threshold values of the corresponding requirement criteria were raised to above the 10th percentile. Specifically, the number of the generated solutions reduced by about 90%, 95% and 100% after increasing the threshold values from the 10th to 90th percentiles at 50%, 75% and 100% satisfaction levels respectively, as illustrated in figure 4.12. At the 100% satisfaction level, the number of generated solutions reached zero at the 70th percentile. This is because of the mutual contradictions inherent in some of the requirement criteria which make their concurrent satisfaction difficult.

Raising the threshold values and satisfaction level of the stakeholder requirements to very high values, therefore, creates risks of losing good alternative solutions at this stage of planning by increasing the effect of the conflicting requirements. This may in turn lead to a sub-optimal rail transit system at a later stage of planning, where the network of lines are generated to connect the candidate stations within the context of an optimization process. This is investigated in chapter 6.

4.5 Verification and Validation of the Algorithm

This section aims to verify and validate the proposed algorithm to ensure that it performs as intended, and to examine whether its structure is appropriate to select “good” candidate solutions. The algorithm verification test was carried out by making the absolute solution for the problem known. It was assumed that there were no areas in the selected case study that are environmentally sensitive and thus the algorithm had to return zero excluded grids due to environmental conditions. The test was conducted with a complete lack of environmental sensitive areas and the algorithm generated the known solution successfully, as illustrated in figure 4.13.

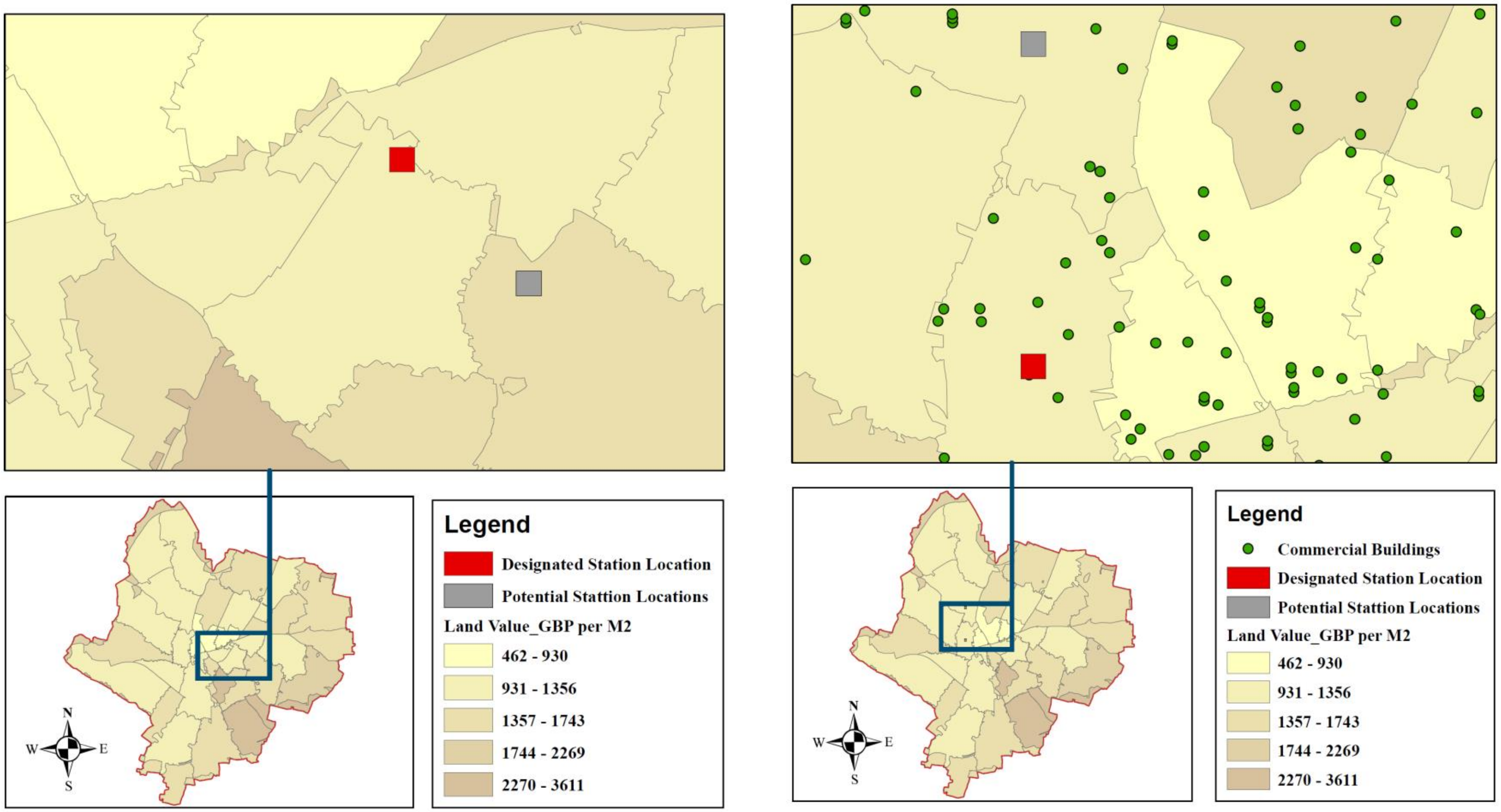


(a) **(b)**

Figure 4.13: Model Verification Using Environmentally Sensitive Areas (ESA): (a) Including ESA, (b) Excluding ESA

The algorithm validation test was implemented using the same notion (i.e. the absolute solution was made known) which was similar to what was proposed by Parker (1977) in validating a model of a rural highway route corridor selection. The same datasets and parameters described in section 4.3 were used for the validation test. Two station grids at the city centre were selected for the test where the assigned corresponding requirement criteria parameters were almost identical to both the grids except for the land values. This test was aimed at analysing the ability of the algorithm to find economically viable solutions that could efficiently satisfy the objectives of the three interested stakeholders. The result shown in figure 4.14a reveals that the algorithm returned the known solution and chose the station grid that had a lower land value.

Another test was conducted to validate the algorithm using the same dataset and parameters described in section 4.3. Similarly, the test was implemented on two station grids, located at the centre and north of the city, as shown in figure 4.14b. The values of the requirement criteria parameters assigned to the grids were equal except for the extent of commercial land use. The result shown in figure 4.14b reveals that the applied algorithm did return the known correct solution and selected the grid at the city centre. This is because the extent of commercial land use around the city centre grid was greater compared to the other grid. The integration of rail transit stations into commercial land use areas is very important to facilitate passenger activity, such as for business, shopping and tourism, enhance operator productivity and satisfy community economic goals. Therefore, it is concluded that the initial results are reliable.



(a) **(b)**

Figure 4.14: Model Validation Using: (a) Land Values (b) Extensiveness of Commercial Buildings

4.6 Summary

This chapter has described a new GIS based algorithm to measure the feasibility of station sites and to determine a candidate pool of feasible station locations. This algorithm aims to help rail transit planners to initially evaluate different station locations based on a comprehensive consideration of various requirements and constraints arising from the different stakeholders in rail transit system planning. Such a tool can effectively speed up the planning process and improve the decision-making capabilities in selecting reliable and cost effective solution alternatives. It can also be viewed as a building block for generating a network of lines linking the candidate stations to obtain the final architecture of the proposed rail system within the context of an integrated optimization process.

The framework of the proposed algorithm consists of five modules which are integrated by exchanging data within an ArcGIS environment. These modules are designed to (1) collect the required planning datasets and rail transit planner preferred planning parameters, (2) evaluate different station locations and determine a pool of feasible stations, and (3) display the customized output of the identified feasible station locations. The algorithm offers the flexibility to be used in different countries/areas by allowing planners to set the planning parameters based on the land use pattern, demography, geography and topology of the study area. In addition, it incorporates flexibilities to be used for both underground/metros and over ground rail transit system planning.

To demonstrate its effectiveness in finding good solutions, the algorithm was applied to a real world case study, the City of Leicester (UK). The results demonstrate that the algorithm can find solutions that satisfy the identified planning requirements. A comparative analysis was also conducted to examine the effects of both the threshold values and satisfaction levels of the rail transit planning stakeholders' requirements on the number of the generated solutions. The results show that raising each of the threshold values and satisfaction levels of the stakeholders' requirements simultaneously was associated with a significant decrease in the number of feasible solutions. This was mainly because of the contradicting nature of some of the requirements. In addition, a set of verification and validation test were conducted and the results demonstrated the ability and robustness of the proposed algorithm to determine "good" feasible solutions.

Chapter 5 details the developments of an integrated optimization algorithm that can simultaneously select the best set of the feasible stations from the solution set identified in this chapter, and optimize the rail line network between these selected stations. It also demonstrates how the algorithm accounts for various local factors and integrates the various planning requirements and constraints to realize a rail transit system that best fits the desired objectives.

Chapter 5 –Optimization of Station Locations and the Associated Line Network

Chapter 4 proposed a GIS based algorithm to identify the potential station locations while accounting for the various requirements and constraints that arise from the rail transit system main stakeholders; passengers, operators and the community. Building on this, this chapter develops a heuristic optimization algorithm to simultaneously select the best set of stations from the station pool and to generate an optimal line network to connect the selected stations. The optimization algorithm, which is based on GA and a background GIS database, is designed to consider various local factors such as travel demand pattern, land use pattern, topography and soil conditions. This is in addition to accommodating complex correlations and interactions of the requirements of the three rail transit system stakeholders, making trade-offs to achieve a reliable and cost effective system.

Section 5.1 outlines the requirements of the passenger, operator and community stakeholders, identified in chapter 3, and presents an overview of the corresponding criteria of these three stakeholders for evaluating different alternative solutions. Section 5.2 presents the framework of the proposed integrated optimization algorithm, which is based on GA and a background GIS database, simultaneously to determine the optimal locations of stations and the associated line network connecting the stations. This section also outlines key steps in the optimization algorithm, which involves the generation of the alternative solutions, fitness evaluation and evolution of the alternative solutions, and determination of the optimal solution. Section 5.3 details how the proposed optimization algorithm generates a set of initial alternative solutions while satisfying the various general requirements of rail transit system planning. Section 5.4 details the fitness evaluation of the alternative solutions generated in section 5.3 with respect to the mathematical functions of the requirement criteria identified in section 5.1, elaborating in detail how the proposed optimization algorithm incorporates various local factors to evaluate alternative solutions via interaction with the GIS supporting system. This section also elaborates in detail the embedded rail transit demand forecasting module and presents its interaction with the optimization algorithm. Section 5.5 details how the proposed GA operators evolve the initial alternative solutions generated in section 5.3 based on their fitness values evaluated in section 5.4

to begin to arrive at an optimal solution. Section 5.6 elaborates the determination of the optimal solution, and section 5.7 summarises the work in this chapter.

5.1 Overview of the Rail Transit Requirements

The main requirements for rail transit system planning are classified into three main groups; passengers, operators and the community, as discussed in chapter 3. Each of these groups has a set of requirements which are strongly affected by the locations of stations and configurations of the line network connecting these stations, as depicted in table 3.1. The satisfaction of some of these requirements mainly depends on the decision of where to place stations, while the satisfaction of others depends also on the final configuration of the line network connecting the stations. Accordingly, the initial locations of stations are determined based on the comprehensive consideration of the various rail transit station related requirements, as demonstrated in chapter 4. Since these stations will be used as a pool from which the best set of stations will be selected, the station location related requirements will not be included in the evaluation framework of this stage. Instead, the requirement sets that are influenced by rail transit system layout will be incorporated into the evaluation framework as shown in table 5.1.

Table 5.1: Requirement Criteria for Determining Rail Transit System Configuration

Stakeholders	Rail Transit System Planning Requirements	Rail Transit System Configuration Related Requirement	Rail Transit System Configuration Related Requirement criteria
passenger	• station availability • travel speed • integration with other transport modes	• travel speed	• train speed, station spacing and line alignment geometry
operator	• station availability • travel speed • integration with other transport modes • population coverage • passenger attraction	• travel speed • population coverage	• train speed, station spacing and line alignment geometry • population density around station areas and station density
community	• station availability • population coverage • passenger attraction • economic growth • environmental impacts • construction cost	• population coverage • construction cost	• population density around station areas • costs associated with building stations and line network.

As illustrated in table 5.1, the requirements that need to be considered when seeking the optimal rail transit system layout are travel speed, station spacing, population coverage and construction cost. This implies that a good rail transit system should be planned so that it provides short travel times to a maximum possible number of people with a minimal construction cost. Achieving maximum coverage of the population with a minimum investment cost is very difficult, however, if it is not impossible. This is because providing a maximum population coverage demands not only a large number of stations along the line network but also placing stations in more developed areas (i.e. those with high density residential or commercial land uses) which are always associated with high land values. This obviously tends to increase the construction cost of the system significantly. A balance between these conflicting requirements is therefore crucial to the acquisition of a reliable and cost-effective rail system. Subsequently, the proposed optimization algorithm is designed to resolve the essential trade-off between maximum rail system usage and minimum passenger travel time on one hand and minimum construction cost of the system on the other hand while also complying with the geometry and operational constraints (defined in section 3.4.2.1 and 3.4.2.2). This translates into the provision of adequate service quality and benefits for passengers and community, as well as economically efficient operation and investment cost. The remaining sections explain the formulation of the optimization algorithm and detail how it seeks to identify the solution that best fits the desired objectives. It should be noted that capturing maximum rail system usage (trip coverage) as opposed to maximum population coverage is more appropriate and realistic, particularly when the optimal layout of the line network needs to be determined, as discussed in section 2.1.1.4. Hence, the maximization of rail system usage is used in the optimization algorithm.

5.2 Framework of The Integrated Optimization Algorithm

Figure 5.1 illustrates the framework of the proposed integrated optimization algorithm which is designed to simultaneously determine the optimal locations of stations and to generate the best line network connecting these stations. As depicted in figure 5.1, a heuristic solution algorithm based on GA is employed to seek a solution that best fits the desired objectives of the three interconnected stakeholders' requirements. A GA is an evolutionary optimization technique widely utilized for solving various complex and large-scale problems. Following the concept of biological evolution, GAs evolve a population of different alternative solutions towards the optimal solution. The evolution begins with a population of random alternative solutions and progresses over a number of generations/iterations. In each generation/iteration, the fitness of the population is evaluated with respect to the desired objectives of the problem. Thereafter, a number of individuals are selected stochastically from the current population based on their fitness values and modified with genetic operators (crossover and mutation) to produce a new population for the next generation/ iteration. This process is continued until a specified number of generations, which is fixed a priori, either have elapsed or other predetermined convergence criteria are met (Goldberg, 1989). The remaining sections detail the key steps of the proposed GA based optimization algorithm which involves generation of alternative solutions, fitness evaluation of the alternative solutions, evolution of the alternative solutions and determination of the optimal solution.

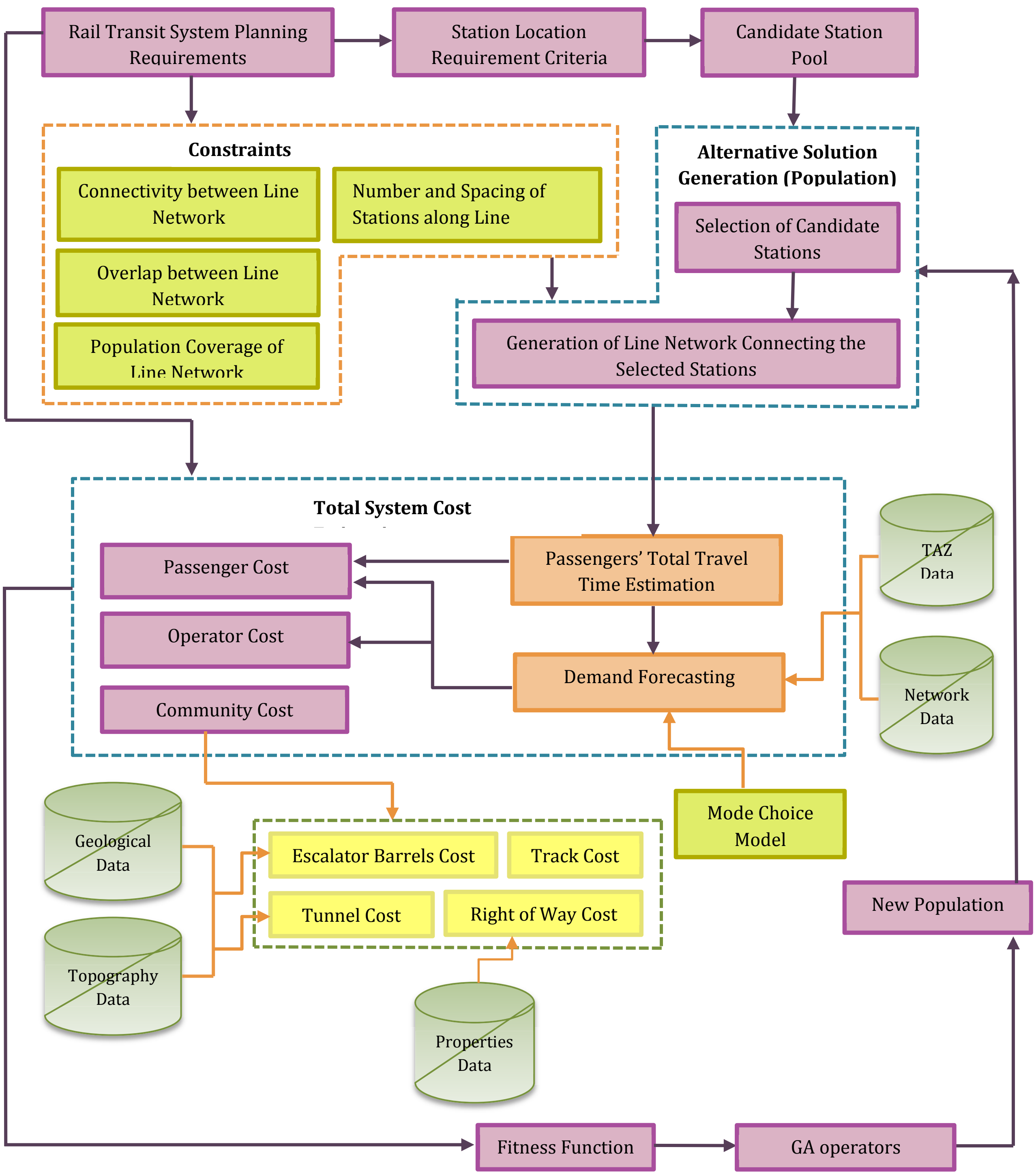
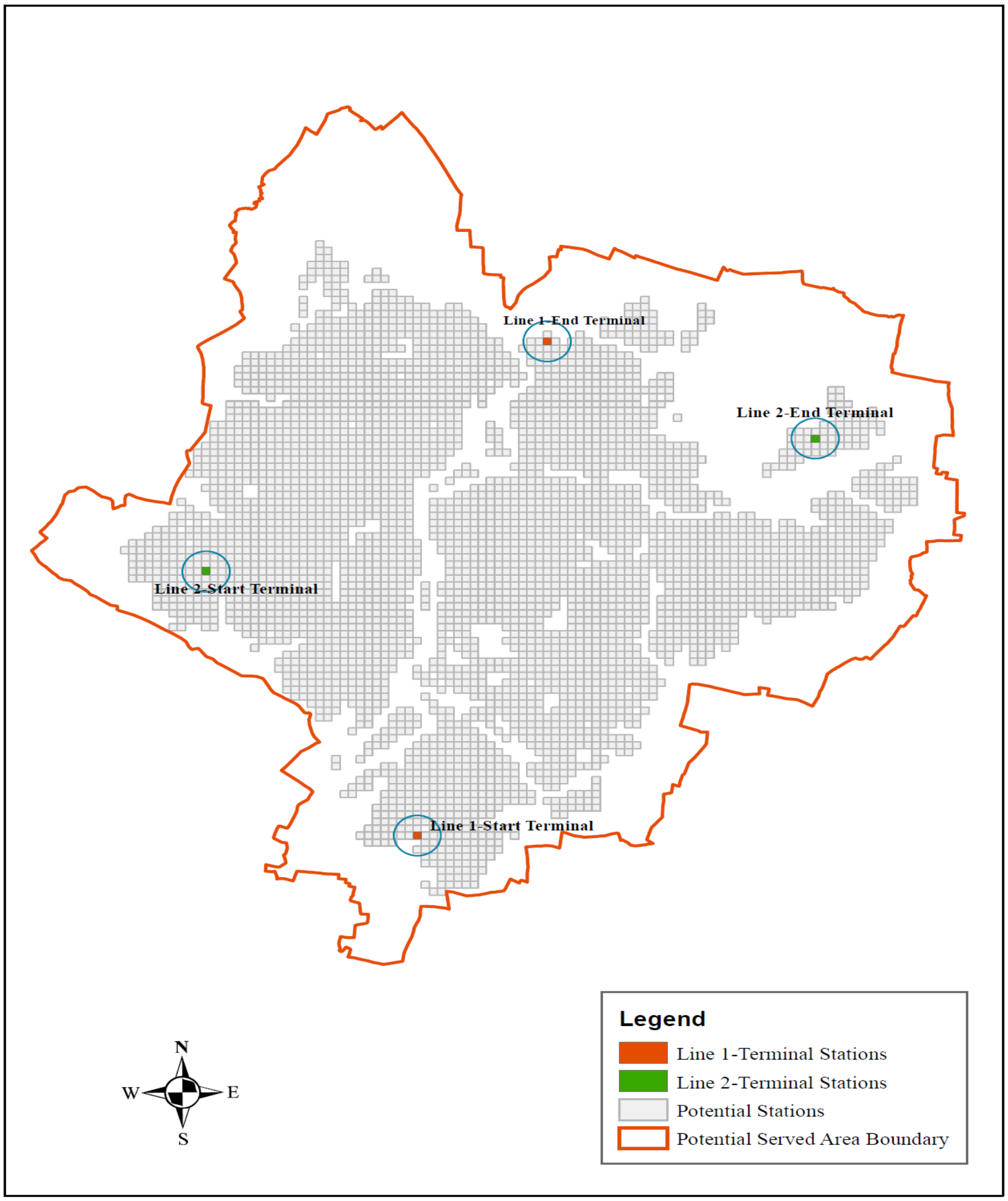


Figure 5.1: Framework of the Integrated Optimization Algorithm

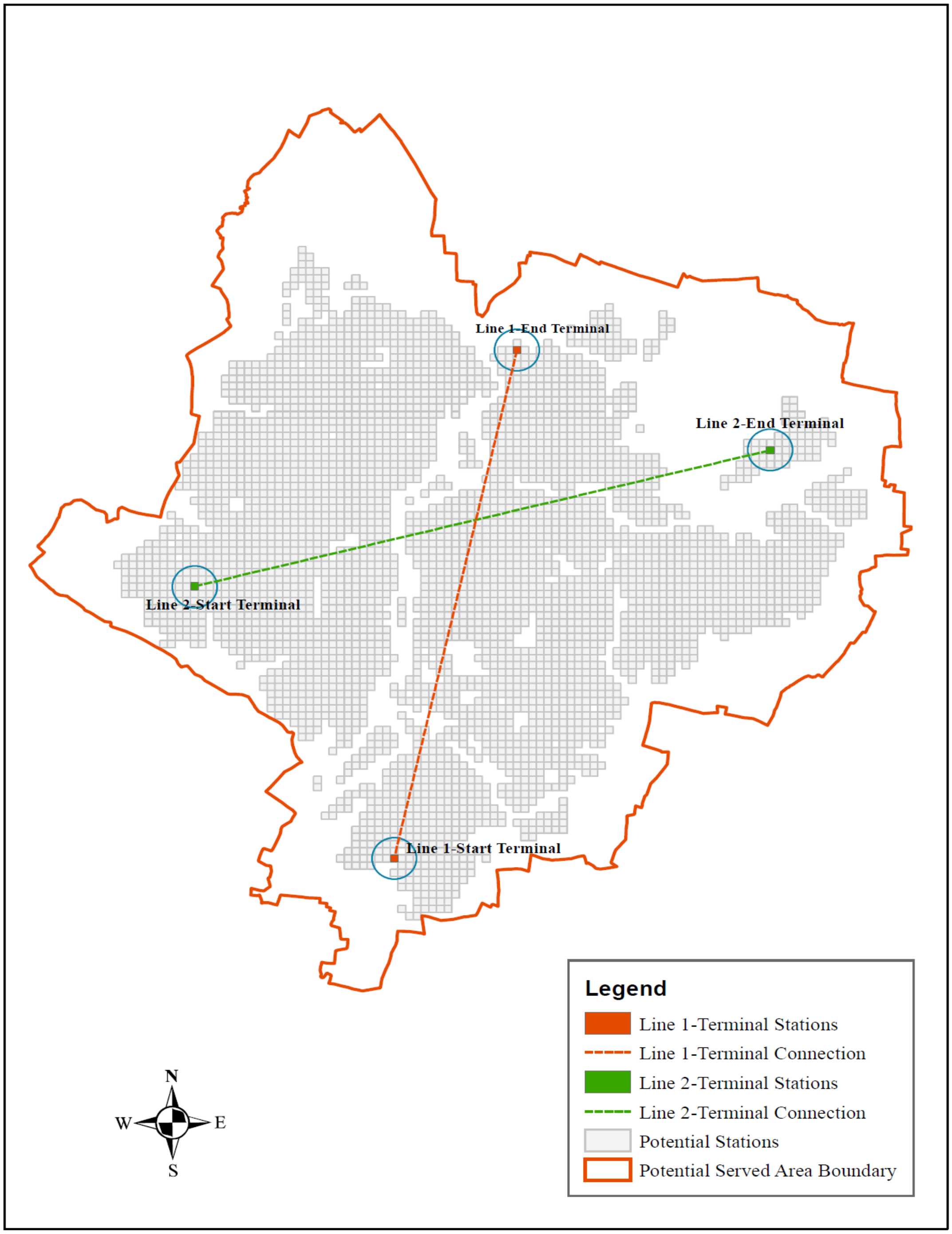
5.3 Generation of the Alternative Solutions

The generation of the alternative solutions consists of selecting a set of stations from the candidate station pool identified in chapter 4 and generating a network of lines connecting the selected stations. It is assumed that, depending on the major traffic flow patterns of the area / city to be served, rail transit planners are able to determine the number of lines desired and the terminal stations of each line (i.e., the beginning and end points of each line, see Figure 5.2a). This allows an initial straight line connection between these two points (Figure 5.2b). Based on the line connection, a corridor is generated around the line with an adjustable width value specified by the planners (Figure 5.2c). This assumption, specifying a priori corridors for line network, is very interesting from both practical and computational perspectives because it: (1) directly integrates rail transit planners' knowledge of the main traffic corridors in a served area/city into the planning platform and therefore replicates what is often done in real world planning practice, (2) limits the location choices for each line in the network which both reduces the computational burden of solving the problem and increases the likelihood of achieving a reliable solution. This explains why this assumption is made in many studies on rail transit system planning and design (e.g., (Bruno et al., 2002, Samanta and Jha, 2008, Gutiérrez-Jarpa et al., 2013). It is also important to note that the basic topological configurations of the alternative solutions are restrained by the pre-specified corridors of their corresponding line network since the line network alignments and associated stations must be located within these corridors.

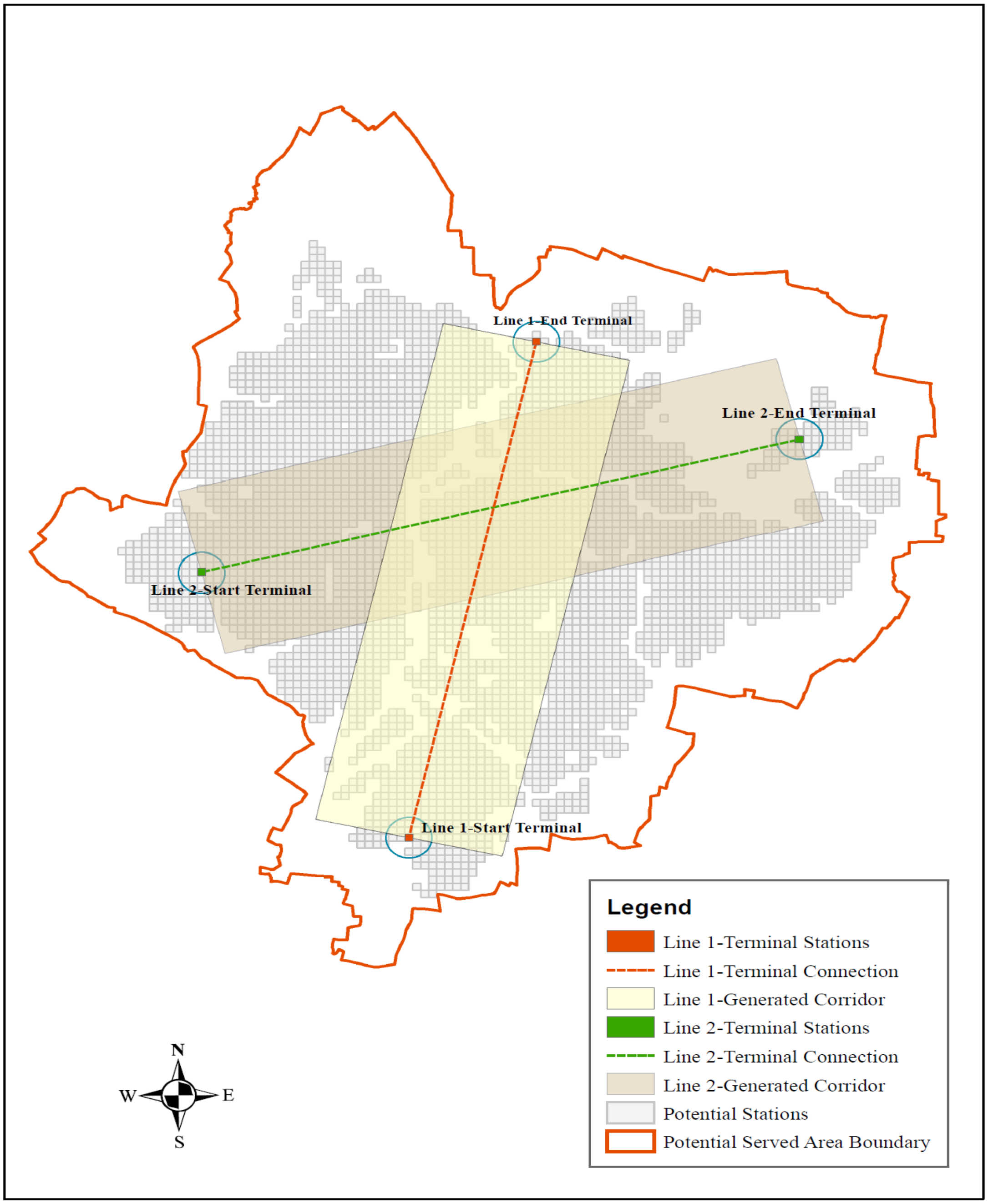
The system topology directly affects passenger service quality, operator productivity and the community benefits, and thus it is very essential that rail transit planners have a good understanding of various topologies and experiences from other cities so as to ensure the effective operation and functionality of the system. The topology of a rail transit system, specifically underground/metro networks, can be classified into several basic types, including star, triangle, cartwheel, and grid, which can be found in many cities. For example, the city of Minsk exhibits a star configuration, the Prague and Kiev metros have a perfect triangle configuration and the Chicago, Montreal and Toronto metros are examples of a modified grid layout. However, the London, Moscow and Paris metros are very complex systems which have combinations of cartwheel and triangle configurations. The basic topological configurations that can be considered in this study are not restricted to a specific type since the proposed optimization algorithm can accommodate any basic topology and find the solution that best fits the desired objectives. However, it is notable that, typically, a rail transit system is built incrementally starting from a simple topology such as a star and a triangle, and with time it evolves into a more complex topology.



(a)


Line 1-End Terminal
Line 2-End Terminal
Line 2-Start Terminal
Line 1-Start Terminal
Legend
Line 1-Terminal Stations
Line 1-Terminal Connection
Line 2-Terminal Stations
Line 2-Terminal Connection
Potential Stations
Potential Served Area Boundary
N
W
E
S

(b)



(c)

Figure 5.2: Generation of Proposed Rail Line Network Corridors (a) Line Network Terminal Stations, (b) Line Network Terminal Connections, (c) Line Network Corridors

The key steps for generating the system alternative solutions are:

- **Step 1:** generate corridors for every terminal station pair corresponding to heavy traffic flows (see figure 5.2).
- **Step 2:** create a viable station list for each terminal station pair from the candidate station pool with respect to the spatial structure of the corresponding generated corridor. That is, the viable station list of each particular terminal station pair encompasses the candidate stations that are laid within the generated corridor of that particular terminal station pair. It is, therefore, possible for each candidate station to be included within more than one viable list according to the configurations of the generated corridors.
- **Step 3:** rank the stations of each viable list with respect to their distance from the corresponding starting terminal station.
- **Step 4:** from the viable station list of each corridor, select the first station departing from the corresponding starting terminal station randomly while satisfying a predefined minimum and maximum station spacing, and generate a line alignment in between them (i.e., the terminal station and the first station subsequent to it).
- **Step 5:** in a similar manner, select the next station departing from the previous selected station, and continue until the end terminal station is reached while ensuring that the number of stations in each corridor / line falls within predefined minimum and maximum limits set by rail transit system planners.
- **Step 6:** check the population coverage of each line generated in step 5, which must be equal to or greater than a predefined threshold value for the line to be accepted as a candidate line. In the case where a particular line of the generated line network does not satisfy this constraint, then the corresponding line alignment and the associated station set are declined, and steps 4 and 5 are repeated to generate a new solution for that particular line. The population coverage of each line in the network is computed using equation (3.17), which divides the total number of people who live within a walking distance of the stations that are located along the line under consideration by the total number of people who live within the walking distance of that particular line alignment.

- **Step 7:** for each of the generated system solutions check whether it satisfies (1) connectivity constraints and (2) overlap constraints. As discussed in chapter 3-section 3.4.2.2, these constraints are checked using equations (3.15) and (3.16). The connectivity constraint ensures that the line networks intersect to ensure easy transfer for passengers between lines. The overlap constraint prevents the occurrence of lengthy overlapping sections between the lines to avoid extending the lines unnecessarily, thereby ensuring that the system has sufficient geographical extent and is available to a maximum possible number of people. In the case where a particular line of the generated line network does not satisfy any o of these two constraints, the corresponding line alignment and the associated station set are declined, and steps 4 and 5 are repeated to generate a new solution for that particular line only.

It is notable that the generated alternative solutions largely depend on the set of terminal station pairs and constraints applied for checking the feasibility of the generated solutions. The planner's knowledge is therefore, vital to the acquisition of an optimal rail transit system.

In GAs, a set of candidate solutions is called a population; each candidate solution in the population pool is known as a chromosome or an individual; each decision variable of the solution (i.e., chromosome) is known as a gene; each parameter of the decision variable (i.e., gene) is known as a bit. Accordingly, the proposed heuristic algorithm treats the alternative solutions as chromosomes, associated lines as genes, and associated stations of the lines as bits. Figure 5.3 illustrates the schematic diagram for the generated alternative solutions.

	Line 1					**Line 2**							**Line n**				
Sol.1	S10	S2	S7	S9	S15	S6	S9	S13	S12	S20	S22	. . . .	S3	S4	S12	S19	S20
Sol.2	S1	S3	S8	S19	S15	S6	S7	S8	S20	S29	S22	. . . .	S3	S10	S29	S23	S20
. . . .																	
Sol. n	S1	S5	S11	S14	S15	S6	S9	S11	S24	S19	S22	. . . .	S3	S6	S13	S24	S20

Figure 5.3: Generated Alternative Solutions

5.4 Fitness Evaluation of the Alternative Solutions

The proposed optimization algorithm evaluates the fitness of each alternative solution in terms of cost. The lower the cost is, the better it fits the desired objectives. Based on the rail transit system planning requirements, identified in chapter three, the fitness value of each alternative function is a function of three main components; passenger, operator and community costs.

$$T_c = \varphi p P_c + \varphi_o O_c + \varphi c C_c \qquad (5.1)$$

Where: T_c is the total system cost, P_c, O_c, C_c are the passenger, operator and community cost respectively over the life span of the proposed rail system, and φp, φo and φc are the coefficients of passenger, operator and community cost respectively.

As discussed in section 3.4, the reason of integrating these coefficients (φp, φo and φc) into the evaluation process is to incorporate flexibilities into the proposed algorithm to allow rail transit planners to prioritize a particular stakeholder requirement over the others when required. The mathematical formulations of each of these cost components are presented in section 3.4.2.1. The following subsections, however, further detail these cost functions and comprehensively explain how they are calculated. Some of the mathematical cost equations (3.10, 3.11, 3.12) that were defined in chapter 3 are therefore, presented again in these subsections as a reminder and to make it easier to understand the sequence of the evaluation process.

5.4.1 Calculation of the Passenger Cost

The passenger cost is defined as the time cost difference between utilizing the proposed rail system and other existing transport modes, specifically buses and cars. Assuming the proposed system will offer services on a daily basis, the annual passenger cost (A_P) is:

$$A_P = 365 \times \sum_{l=1}^{q} \left(\sum_t \sum_i \sum_j \left((P_{ij} \times T_{Cttij} - P_{cij} \times T_{Ctcij}) + (P_{ij} \times T_{Cttij} - P_{bij} \times T_{Ctbij}) \right) \right)$$

Where:

q is the number of lines in the network;

P_{ij} is the number of railway passengers from zone i to zone j in time period t;

T_{Cttij} is the cost (£) of total travel time by train from zone i to zone j in time period t;

P_{cij} is the number of railway passengers from zone i to zone j in time period t that may switch to car if the rail system is not available;

T_{Ctcij} is the cost (£) of total travel time by car from zone i to zone j in time period t;

P_{bij} is the number of railway passengers from zone i to zone j in time period t that may switch to bus if the rail system is not available;

T_{Ctbij} is the cost (£) of total travel time by bus from zone i to zone j in time period t.

The total train time (T_{Ctt}) includes passenger's access time to/from station, waiting time at station and on-train travel time. Thus, (T_{Ctt}) is calculated as follow:

$$T_{Ctt} = T_{ta} \times A_0 + T_{tw} \times W_0 + T_{tt} \times T_0 \quad (5.2)$$

Where:

T_{ta} is the passenger's access time to / from station (min);

A_0 is the unit costs of access time (£/min);

T_{tw} is waiting time at station (min);

T_{tt} is on-train travel time (min) and;

T_0 is the unit cost of on train/ in-car travel / on- bus time (£/min).

The passenger access time to/from rail stations is a function of the passenger's walking distance and speed to/from stations. It is computed through artificial links created between the traffic analysis zones (TAZs) of the study area and potential station locations. These links measure the distance between the centroids of each TAZ and potential station location (D_a) , thus the access time is calculated by dividing this distance (D_a) by the passenger walking distance (V_a) .

$$T_{ta} = \frac{D_a}{V_a} \quad (5.3)$$

The waiting time depends on the frequency of trains, the reliability of the scheduled train frequency and passengers' arrival pattern at the boarding station, as discussed in section 2.1.1.1. Since none of these elements can be evaluated accurately at this stage of the planning process, the waiting time is assumed to be equal to a half of the train headway (H_w), which is a function of train frequency. The changes in train frequency and passengers' arrival pattern at different time periods due to changes in the congestion level is also accounted for by incorporating flexibility into the algorithm to allow planners to set different values of (H_w) for different time periods.

$$T_{tw} = \frac{1}{2} \times \; H_w \tag{5.4}$$

The on-train travel time (T_{tt}) depends on the distance between boarding and alighting stations (D_s), the speed of the train (V_t), and the dwell time at stations in between the stations where passengers boarded and alighted their train (n_s), which can be represented as a fixed value (T_d). Thus, T_{tt} is calculated as:

$$T_{tt} = \frac{D_s}{V_t} + (n_s \times T_d) \tag{5.5}$$

The travel distance between each boarding and alighting station in the network (D_s) is computed by implementing the Dijkstra's shortest path algorithm, which is embedded into the evaluation framework of the proposed optimization algorithm. This algorithm treats the proposed network as a graph, in which the stations represent the nodes of the graph and the lines linking the stations represent the edges of the graph. With a defined non-negative weight of the edges, which can be cost, length, or travel time, the algorithm seeks the shortest path for every node-to-node trip as a set of sequenced nodes or links. As a result, minimum cost, travel distance or travel time paths are generated for all node pairs in the graph. The details of this algorithm are presented in (Johnson, 1973, Knutm, 1977, Chen, 2003). In this study, the weight assigned to the lines linking the selected candidate stations is defined as their corresponding length while the algorithm searches for the shortest path between each boarding and alighting station pair, and thus the minimum travel distance between each boarding and alighting station (D_s) is obtained.

Unlike the total train time, the total car time (T_{Ctc}) consists only of in-car travel time between origin and destination and the time spent searching for a parking space, and thus (T_{Ctc}) is calculated as:

$$T_{Ctc} = T_{ct} \times T_0 + T_{cs} \times S_0 \qquad (5.6)$$

Where:

T_{ct} is in-car travel time (min);

T_{cs} is search time for a park space (min) and;

S_0 is the unit cost of search time for a parking space (£/min).

Similar to the total train time, the total bus time (T_{Ctb}) consists of passenger's access time to/from bus stop, the waiting time at the bus stop and the on-bus travel time.

$$T_{Ctb} = T_{ba} \times A_0 + T_{bw} \times W_0 + T_{bt} \times T_0 \qquad (5.7)$$

Where:

T_{ba} is the passenger's access time to/from bus stop (min);

T_{bw} is waiting time at bus stop (min) and;

T_{bt} is on-bus travel time (min).

The road and bus network travel time information etween each traffic analysis zone pair are assumed to be obtainable from relevant transport agencies for the AM/PM peak, intermediate peak IP and off peak OP periods, thus taking into account traffic congestion levels for different time periods.

The rail transit network usage and its distribution pattern over the existing transport modes (i.e. car and bus) in cases where rail is not available are predicted with a discrete mode choice model, widely used in practice. This model is built into the evaluation framework of the proposed integrated optimization algorithm. Section 5.4.4 describes and explains this model in detail.

Since the passenger cost recurs throughout the life span of the system, a life cycle analysis is performed to calculate the passenger cost of the system over its design life. Thus, assuming a fixed annual interest rate, i, and a design life of n years, the passenger cost over the system life cycle (P_C) is:

$$P_C = A_P \frac{(1+i)^{n-1}}{i(1+i)^n} \tag{5.8}$$

5.4.2 Calculation of the Operator Cost

The operator cost is defined as the difference in the operation and maintenance costs between utilizing the proposed rail system and other transport modes (specifically car and bus), which is derived as a function of each mode's corresponding travelled distance and potential passengers. Assuming the proposed rail system offers services on a daily basis, the annual operator cost (A_O) is:

$$A_O = 365 \times \sum_{l=1}^{q} \Big(\sum_t \sum_i \sum_j \big((P_{ij} \times L_{tij} \times M_{t0} - P_{cij} \times L_{cij} \times M_{c0}) + (P_{ij} \times L_{tij} \times M_{t0} - P_{bij} \times L_{bij} \times M_{b0}) \big) \Big)$$

Where:

L_{tij} is the train km travelled from station i to station j in time period t;

M_{t0} is the unit operation and maintenance cost for train (£/ passenger-km);

L_{cij} is the car km travelled from zone i to zone j in time period t;

M_{c0} is the unit operation and maintenance for car (£/ passenger-km);

L_{bij} is the bus km travelled from zone i to zone j in time period t and;

M_{b0} is the unit operation and maintenance cost for bus (£/ passenger-km).

The train km travelled for every station pair in the proposed network is calculated using the Dijkstra's shortest path algorithm, as noted in the previous section. The bus and car travel distance between each traffic analysis zone pair are assumed to be obtainable from relevant transport agencies at different time periods (i.e., AM, IP, OP and PM). The demand values are forecasted using a discrete mode choice model, which is incorporated in the proposed optimization algorithm framework (see section 5.4.4).

Similar to the passenger cost, due to the recurring cost throughout the lifespan of the system, the operator cost can be calculated for the whole system life (O_C) as follows:

$$O_C = A_O \frac{(1+i)^{n-1}}{i(1+i)^n} \tag{5.9}$$

5.4.3 Calculation of the Community Cost

The community cost is defined as the required capital cost for constructing stations and the line network connecting the stations, which is classified into location-dependent and size-dependent cost categories. The former category involves those construction items that are influenced by the different locations of the stations and associated line network. Therefore, the category covers the costs for acquiring land, building tunnels and escalator barrels, which are affected by land use density, soil conditions and topography. The size-dependent cost category involves those construction items that are proportional to the length of the line network and size of the associated stations. The associated costs for building and equipping stations with the required facilities, laying rail tracks and fastening rails are therefore classified as the size-dependent costs. The community cost ($\mathrm{C_c}$) is therefore, calculated by summing up the location- dependent (C_{LDep}) and size-dependent (C_{SDep}) costs.

$$\mathrm{C_c} = C_{LDep} + C_{SDep}$$

$$C_{LDep} = \sum_{l=1}^{q}(Rw_l + Tu_l + Es_l) \tag{5.10}$$

$$C_{SDep} = \sum_{l=1}^{q}(Sb_l + Tr_l) \tag{5.11}$$

Where:

q is the number of lines in the proposed rail network system,

Rw_l is the right of way (i.e., land acquisition cost) for stations along line *l* (£),

Tu_l is the cost of building the tunnel structure along line *l* (£),

Es_l is the cost of building station escalator barrels along line *l* (£),

Sb_l is the cost of building and equipping the stations with necessary facilities along line l (£),

Tr_l is the total cost of laying rail tracks and fastening rails along line l (£).

These two cost categories may favour different line configurations due to their different behaviours. The location-dependent cost items tend to favour crooked and circuitous lines, while the size-dependent cost items tend to favour short and straight lines. The optimization algorithm, therefore, tends to make a trade-off between these different cost types while seeking the optimal solution. Even though, these construction cost parameters are set for underground rail transit systems. They can be modified for over-ground rail transit system construction parameters by including those construction parameters that may incur if the system is built over ground, such as cost of building bridges.

It should be noted that the community cost over the system lifespan is not calculated here. This is due to the fact that, unlike the passenger and operator costs, the construction cost is incurred once (i.e., at the onset of the project).

Since the computation procedures for the aforementioned construction cost items are different from each other, they are explained in five separate subsections. The first three subsections elaborate the estimation of the location-dependent costs, and the last two explain the estimation of the size-dependent costs.

5.4.3.1 Right of Way Cost

The right of way cost includes land acquisition and property damage cost items for building stations and the associated line network. It is a function of where the stations are built and the associated line network constructed. As this thesis focuses on rapid rail transit system/metro, the cost of acquiring right of way depends only on the locations of the stations as the deep tunnels through which the lines connecting the stations are passed entail no right of way cost. This construction cost item is computed via interaction with the GIS, in which the unit cost database of land parcels and properties are stored. The proposed algorithm first identifies all sections of land parcels and properties that are located within the stations' right of way boundaries, and then extracts the corresponding costs of these sections from the GIS database to compute the right of way acquisition cost. Obviously, the right of way acquisition cost is the sum of the cost values of these land parcels and property sections.

5.4.3.2 Tunnel Cost

Tunnel cost usually comprises the largest share of the total construction cost of an underground rail transit system, and directly affects the location and configuration of the system. It comprises costs of tunnel excavations, including the tunnel support system, and equipping the tunnel with the necessary facilities such as traffic control, lighting, fire, ventilation, surveillance and safety systems. Since the former cost item dominates the other cost items, however, and is highly influenced by locations of lines, this study focuses on this item (i.e., tunnel excavation cost).

A reasonable tunnel excavation cost estimate is, therefore, very essential and is required to justify proceeding with the evaluation process. It is well known that the excavation cost of a tunnel is a function of numerous variables which makes its estimation complex. The main factors that affect the excavation cost of a tunnel are tunnel length, tunnel diameter, geological conditions, height of overburden/tunnel depth, geography, ground support and excavation method, which itself is mainly a function of geologic conditions and tunnel sizes and orientations. Also, many non-technical factors, such as availability of a skilled and experienced workforce and contractors can significantly affect the tunnel construction cost. At the early stage of planning, little information on these factors is available to planners, i.e., very limited information is available on the geological conditions and the line alignments are not known yet, the depth is not chosen and even the length of the tunnel is not known exactly. Developing an accurate cost estimate is therefore very difficult at this stage and therefore, a very limited number of cost models are at the engineers' disposal for this purpose. Most of the studies on the rail transit system and highway planning, such as Kim (2001), Kim et al. (2005), Samanta (2005), Lai (2012) and Lai and Schonfeld (2012) have therefore, assumed that the tunnel excavation cost is a linear function of tunnel length and diameter and simply ignored consideration of the other factors. However, ignoring other critical related factors in particular geology and excavation method, carries a risk of yielding grossly inaccurate cost predictions. Clearly, an exact estimate of tunnel cost at the planning stage is not possible, yet it remains essential for the estimated cost to be reliable to allow a valid comparison between alternative solutions and to perform reliable "what if" scenarios relative to the tunnel size and length. Also, the cost estimate generated at this stage is often treated as a basis for all the subsequent design and construction stages and is referred as the baseline cost for the project, and thus there is a great need for a reasonably reliable cost estimate.

Given the issues above, this chapter, develops a tunnel cost estimation model that can be used for various applications in the planning stage based on statistical analysis of historical cost data while taking into account tunnel size and length, geological conditions and excavation methods. In addition, it proposes an algorithm to determine optimal tunnel depth along a rail transit line network and integrates this into the tunnel cost estimate model to compute the total construction cost of the tunnel structures. It is worth mentioning that developing such an algorithm combining the determination of optimal tunnel depth with a tunnel cost evaluation framework for a rail transit line network represents a novel contribution of this research. The detailed development of the tunnel cost estimate model and depth algorithm, and their integration into the evaluation framework of the proposed optimization algorithm, are presented below.

5.4.3.2.1 Tunnel Cost Estimation Model

In general, methods for estimating tunnel construction cost can be classified into three main groups; deterministic, interval and percentile and probabilistic methods (Špačková et al., 2013). The deterministic methods provide a tunnel cost in a single value representing the best or ideal estimate and neglect the uncertainty of the cost estimate, which may occur due to unforeseen geological conditions, variability of the construction performance and unexpected tunnel collapses and settlements. The interval and percentile estimates consider the uncertainty of the estimate by expressing a cost estimate with a confidence interval or a percentile, i.e., linking the cost value to the probability of it being exceeded. The accuracy level of this estimate or the width of the interval, which are determined mainly based on expert judgment, varies at different stages of the project. At the planning stage, because of very limited information on the geological conditions, tunnel design and construction performance and technology, the estimate is unlikely to be very accurate and therefore the interval will be wide. At the detailed design stage, once extensive geotechnical investigations have been conducted, a detailed tunnel design has been prepared and the type of excavation method has been selected, a more precise estimate can be made. The accuracy of the estimate at this stage is therefore high and the interval is less compared to the planning stage.

The probabilistic methods recognize the importance of performing a risk analysis to evaluate and quantify the uncertainty in a cost estimate and, thus, they express a tunnel cost as a probability distribution curve over a range of cost values. In practice, deterministic methods are the most commonly used, sometimes supplemented by a qualitative or semi-quantitative analysis of the cost uncertainties. Although, estimation of the tunnel cost on a probabilistic basis tends to be more accurate and is associated with more robust decisions, it requires detailed geological and hydrological investigations to be carried out to quantify the uncertainties and risks. Since obtaining such detailed information during the planning stage is very difficult, this thesis develops a tunnel cost estimate model on the deterministic basis, which should be sufficient for comparing different alternative solutions and running “what if” scenarios on the tunnel size and extent.

The cost estimation model is developed by establishing a database of tunnel costs from different projects and subsequent regression analysis to determine the trends and proper formulas. Data from 38 tunnel projects were obtained from the British Tunnel Society (BTS) and compiled into a database. The projects were executed in the UK and Western European countries, including Germany, France, Switzerland, Spain, Norway, Austria and Greece, since 2000. The data for each project included recorded construction cost, tunnel length and diameter, ground conditions, location, excavation method and tunnel application. The completion dates of the projects were different from one another and the European projects were in a number of different currencies. The BTS therefore, adopted foreign exchange rates to bring the monetary terms of these projects into a consistent cost baseline, which was GBP. This is in addition to the consideration of construction inflation for each country by applying Retail Price and Building Tender Price indices to bring the compiled data to a specific baseline date. The baseline date was set at Q1 2010. In this analysis, however, the reference date was set at Q2 2014 and accordingly all the monetary terms were converted to this date using Construction Tender Price indices.

In a subsequent step a preliminary analysis was carried out on the information compiled in the database, but the results were not conclusive. This may be due to the fact that each project is unique in its purpose, requirements and environmental circumstances. Therefore, to recognise these differences and achieve better project results the following categorization are applied in this thesis:

1- Ground conditions: based on the ground conditions through which the tunnels were bored, the data were classified into two categories:
    - Rock
    - Soil
2- Tunnelling methods: based on their tunnelling method (excavation and ground supporting system), the data were classified into two categories:
    - Mechanized tunnelling methods which include both Tunnel Boring Machines (TBMs) and shielded machines.
    - Other/Conventional tunnelling methods, which include all the other tunnelling methods apart from mechanized tunnelling methods, such as drilling and blasting, road header, Austrian Tunnelling Method (NATM), and the Sequential Excavation Method (SEM).
3- Tunnel applications: the data were also classified based on their end use for:
    - Highway and Rail
    - Water
    - Power

It should be noted that both the type of ground condition and the type of tunnelling method incorporate the geologic conditions, which are critical factors that affect the tunnel construction process in terms of both cost and time. It is also important to note that both the rail and highway tunnels are merged into one category since they share similarities in many aspects. In addition, their corresponding available data points are small and thus disaggregating them into two categories would leave an insufficient sample for the regression analysis. Figure 5.4 illustrates the pie charts for the above three categories available in the database.

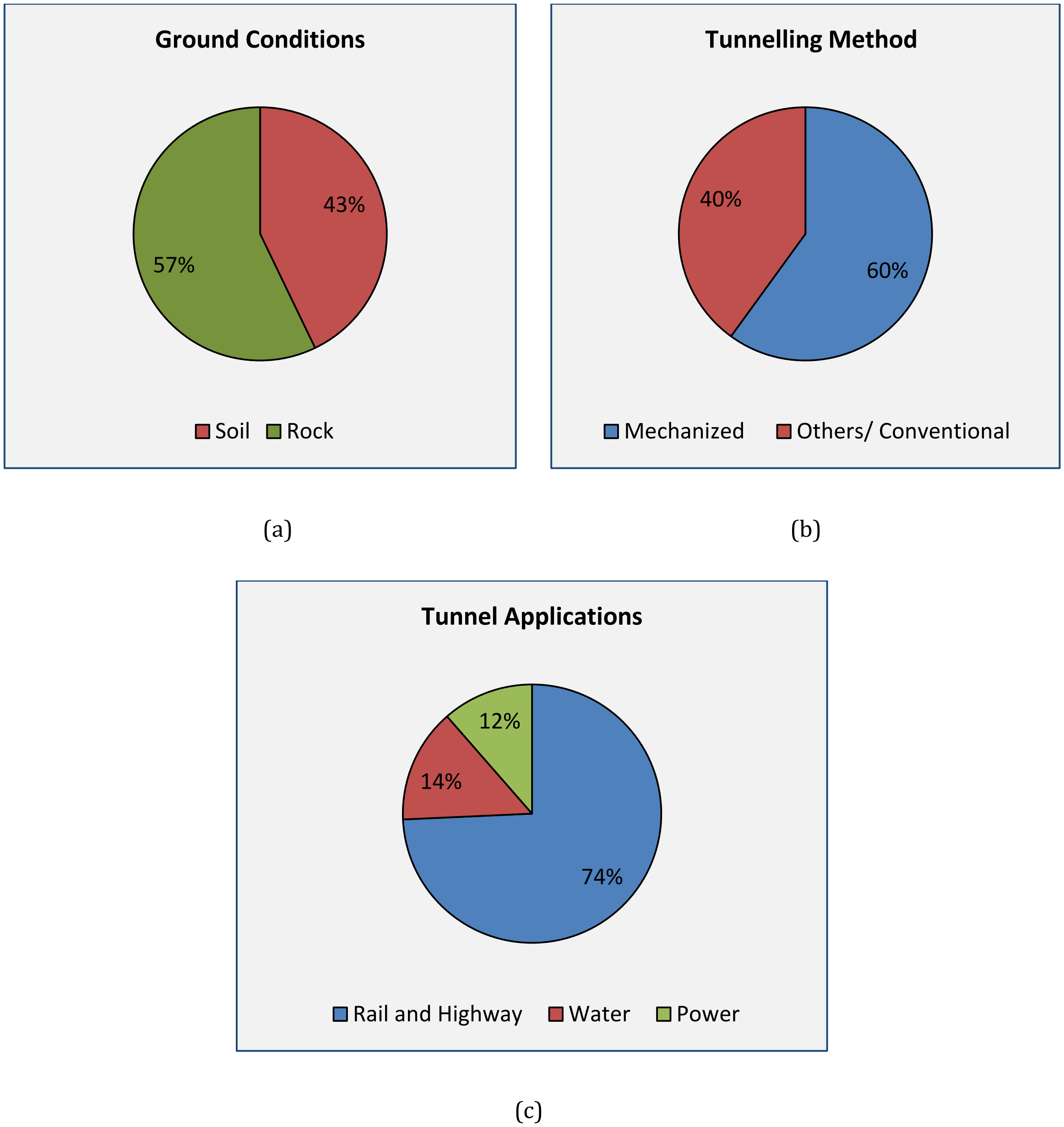


Figure 5.4: Pie Charts of the Tunnel Database Categorization (a) Ground Conditions, (b) Tunnelling Methods, (c) Tunnel Applications

Regression analysis was carried out on the data and the correlation coefficient was computed for highway and rail tunnel types only. Although a linear regression fits the data well for most cases, the best fit was obtained when the log10 of parameters were used. These results point to the fact that the relationship between the explanatory variables (tunnel length and diameter) and the response variable (i.e., tunnel cost) is log-linear. Often log transformation of the variables can "straighten" a nonlinear relationship and thus makes the variables to fit better linear regression than their original raw values.

The results of the regression analyses for each case are shown in table 5.2. However, the regression analyses for the water and power tunnels were not performed because it is beyond the scope of this research. Also, it is notable that no model was proposed for estimating the cost of tunnelling in soil using the conventional methods due to the lack of available data points. In a subsequent step, a test was applied to each of the proposed cost estimate models to measure their accuracy. This test calculates the ratio of cost estimated from the proposed cost model (Ec) to actual cost (Ac). When calculated, the test value (Ec/Ac) is equal to one, indicating that the cost model can predict with up to 100% accuracy. That is, the closer the test value is to 1.0 the more accurate the model prediction is. It is worth noting that Rostami et al. (2013) applied the same test to measure the accuracy of a set of models developed for estimating the construction cost of various tunnel types including highway, subway, water and waste water tunnels. The result of the accuracy test for each case is shown in table 5.2.

Table 5.2: Summary of Unit Cost and Regression Analyses for Soil and Rock Ground Conditions for Rail and Highway Tunnels

Ground Condition	Tunnelling Method	No. of Data	Ec/Ac	R^2	Regression Equations-2Q 2014
Soil	Mechanized	8	0.851	0.921	Cost(M£)=10^(1.052+1.05Log(L*)+0.467Log(D**))
	Conventional	0	--	--	--------
Rock	Mechanized	7	0.909	0.830	Cost(M£)=10^(1.389+0.778Log(L*)+0.464Log(D**))
	Conventional	10	0.815	0.698	Cost(M£)=10^(-0.537+0.856Log(L*)+1.992Log(D**))

*L is tunnel length in (km), and **D is tunnel diameter in (m)

Figures 5.5 to 5.7 validate the accuracy of the predicted cost in terms of the actual cost for rail and highway tunnels in soil and rock ground conditions. The slope values show that the proposed cost estimate models possess reasonable accuracies of 0.851, 0.909 and 0.815 for mechanized soil and rock ground conditions, and for conventional rock ground conditions, respectively. In other words, there is no significant difference between the slopes of the cost estimate models and the ideal predictors (slope of 1). In addition, these results indicate that the cost estimate models tend to underestimate the cost of very large volume projects by about 10 to 20 % depending on the ground conditions and tunnelling types. Figures 5.5 to 5.7 show that, compared to the tunnelling cost model for conventional techniques in rock ground conditions, the mechanized techniques in both soil and rock ground conditions are more accurately predicted for both small and large tunnelling projects. This is not to suggest that these models are not appropriate for use during the planning stage, but to make planners cautious about the accuracy of these models at the detailed design and construction stages where the project budget is refined and responsiveness of contract bids is evaluated.

The accuracy of the resulting models can be further improved by adding new projects to the database and reanalysing the data to update the models and related formulas. Enlarging the database, however, will still not result in very high prediction accuracies since, as noted earlier, the predicted cost at the planning stage cannot be very accurate as the actual construction cost is largely influenced by the geological conditions encountered during construction, by the tunnel design and by the contractors experience and practice. Applying contingency factors to the predicted cost is therefore, strongly recommended so as to mitigate the possible inaccuracy in the predicted cost that may occur due to unforeseen ground conditions, possible design changes, a lack of experienced contractors and skilled work force and other construction performance related factors. The values of the contingency factor should be based on expert judgment and analysis of projects constructed in the past.

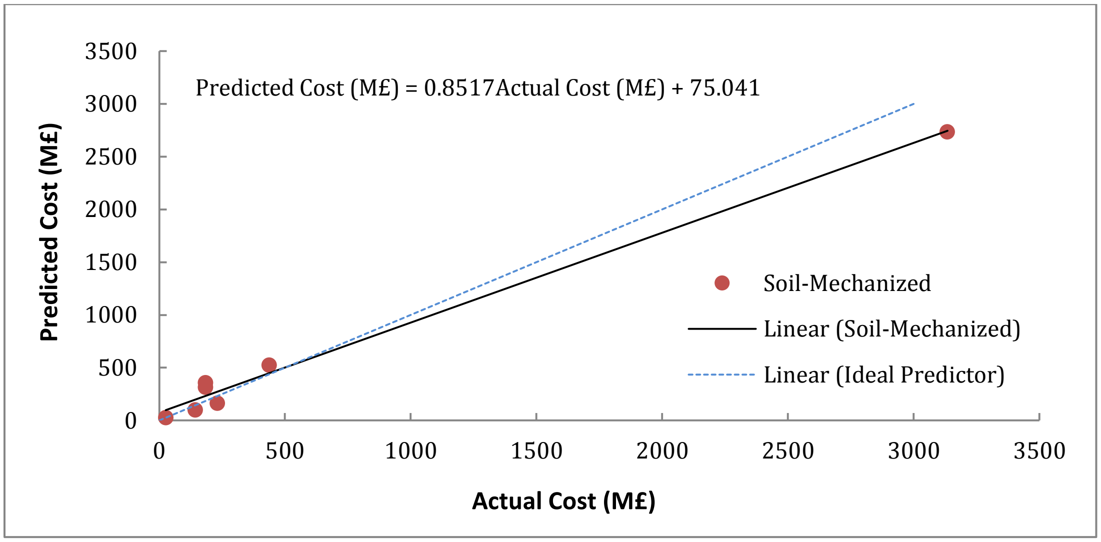


Figure 5.5: Predicted Cost versus Actual Cost for Rail and Highway Tunnels Using Mechanized Tunnelling Methods in Soft Grounds (Soil)

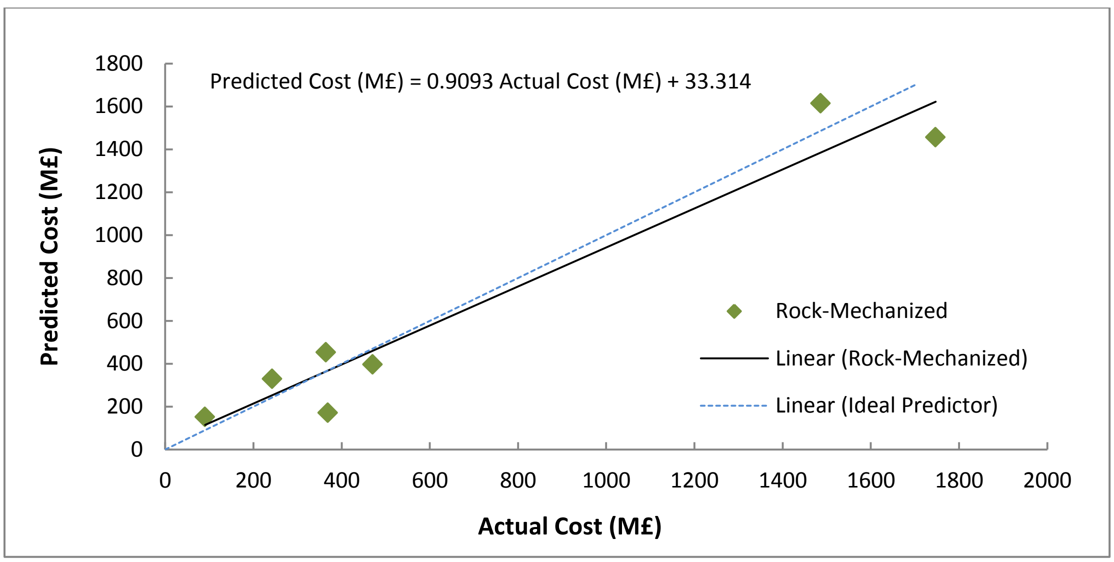


Figure 5.6: Predicted Cost versus Actual Cost for Rail and Highway Tunnels Using Mechanized Tunnelling Methods in Rocks

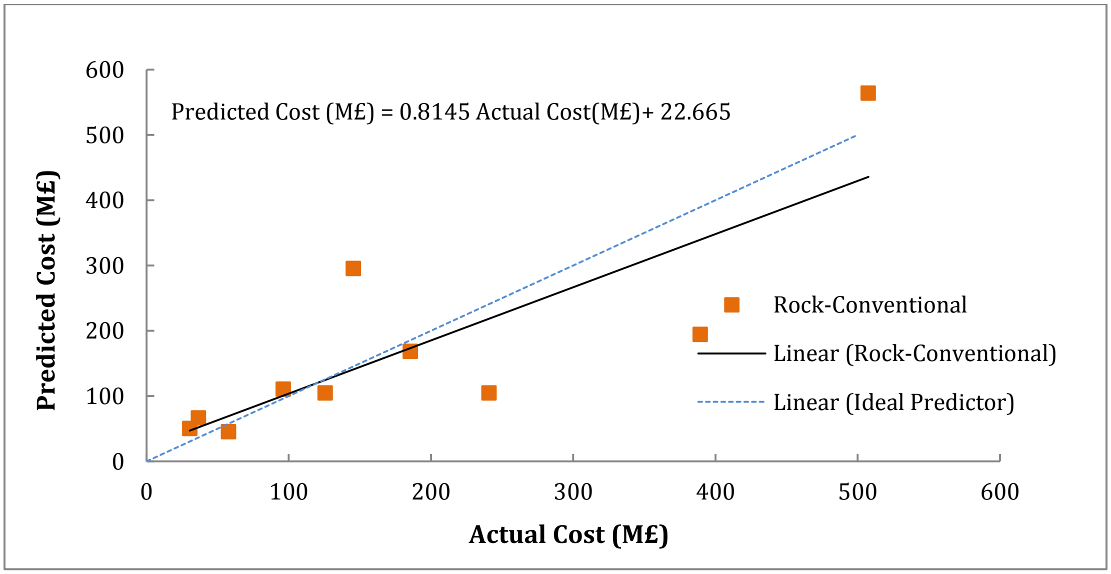


Figure 5.7: Predicted Cost versus Actual Cost for Rail and Highway Tunnels Using Conventional Tunnelling Methods in Rocks

It should be noted that the models proposed in this chapter have been compared to some other developed models for estimating tunnel construction cost at the planning level, specifically models developed by Rostami et al. (2013). The comparisons indicate that although the structure of models in both studies take a similar form, their predictions are different when applied to the same parameters (i.e., the same tunnel diameter and length). These differences can be explained by several factors, but main ones being the size and complexity of the projects analysed in both studies, the different size and source of the compiled data and different geographical locations of the analysed projects. The projects analysed in this thesis were implemented either in the UK or a Western European country, while the projects in the other study were implemented either in North America or Canada. The geographical location of a project has a significant effect on the available workforce pool and specialized skills, which in turn influences the construction process in terms of both cost and time.

5.4.3.2.2 Tunnel Depth Determination

Tunnel depth, or overburden height, is one of the key factors influencing tunnel construction cost, as discussed in the previous section. Basically, increasing the depth of a tunnel is strongly associated with increased uncertainties and the probability of encountering adverse and unexpected ground conditions, which require extensive site investigations in order to be mitigated, resulting in significantly increased tunnel construction costs. Consequently, this chapter develops a heuristic algorithm to minimize tunnel depth along a rail transit line network while considering the intersections of the lines and geometric requirements and incorporating the computed depth into the proposed optimization evaluation framework. The algorithm applies the following key steps while interacting with the GIS supporting system to determine the optimal tunnel depth:

- **Step 1:** extract the ground elevation (GE) of each station in the proposed network generated in section 5.3 from the topography map of the potential area/city to be served using the GIS "extract value" to point to the geoprocessing tool. The topography map must be available in the standard GIS format.
- **Step 2:** for each line in the proposed network, rank the corresponding stations with respect to their ground elevation in ascending order, and set the station that has the lowest ground elevation as St. Base.
- **Step 3:** rank the line network in an ascending order with respect to the station that has the lowest ground elevation (i.e., their St. Base), and append them to a list. This means that the line that has a station with the lowest ground elevation among the line network has a zero index in the list.
- **Step 4:** determine the tunnel depth at each station along the first line in the list generated in step 3 starting with its corresponding St. Base station. The tunnel depth at the St. Base is set to be equal to the predefined minimum tunnel depth (TD min), which usually lies between 15-20 metres below ground for a rapid rail transit system/metro if its type of construction is deep tunnelling. The tunnel depth of the remaining stations along the line is determined based on the tunnel depth of the base station (i.e., St. Base) while considering the required minimum and maximum longitudinal grade of the rail transit line alignments and the required minimum depth from the ground surface (i.e., TD min). Figure 5.8 illustrates this step which is mapped to the following equations to compute the depth of the tunnel at every station located along the line network.

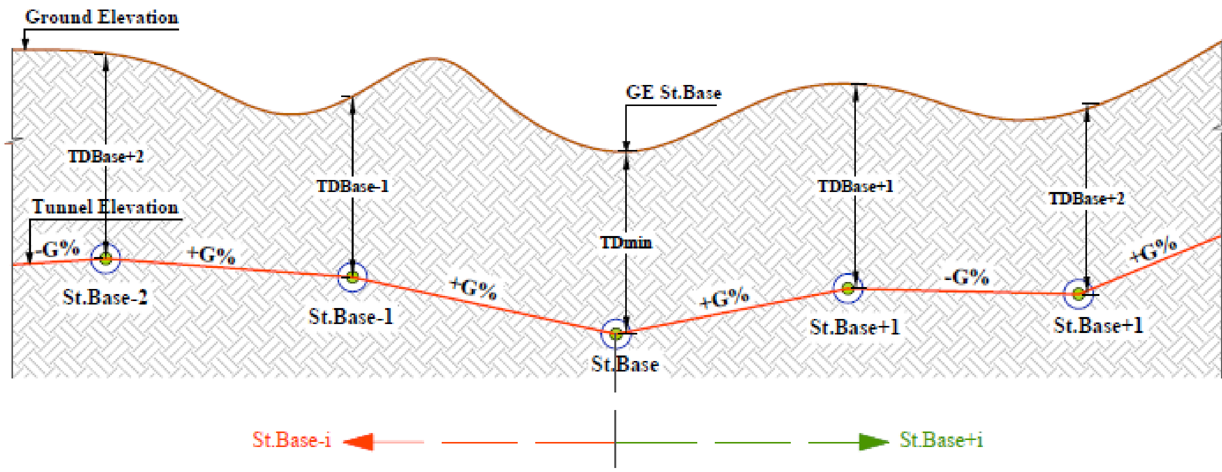


Figure 5.8: Illustration of the Tunnel Depth Calculation

$$TD_{st.base} = TD_{min} \tag{5.12}$$

$$TE_{st.base} = GE_{base} - TD_{st.base} \tag{5.13}$$

$$TD_{St.Base-i} = GE_{St.Base-i} - \left(TE_{(St.Base-i)-1} \mp G \times Dis_{((St.Base-i),(St.Base-i)-1)}\right) \geq TD_{min} \ \& \ G_{max} \leq G \geq G_{min} \tag{5.14}$$

$$TD_{St.Base+i} = GE_{St.Base+i} - \left(TE_{(St.Base+i)-1} \mp G \times Dis_{((St.Base+i),(St.Base+i)-1)}\right) \geq TD_{min} \ \& \ G_{max} \leq G \geq G_{min} \tag{5.15}$$

Where:

TD is the depth of the tunnel,

TE is the elevation of the tunnel,

G is the longitudinal slope of the rail line,

Dis is the distance between stations,

G_{max} *and* G_{min} is the required maximum and minimum longitudinal slope of the rail line,

$St.Base - i$ are stations located before the base station along the rail line under the consideration,

$St.Base + i$ are stations located after the base station along the rail line under the consideration.

The proposed algorithm changes the values of G between G_{max} and G_{min} systematically upwards and downwards to ensure that the depth of the tunnel remains minimal while ensuring that the required minimum depth from the ground surface (i.e., TD min) is satisfied. It is assumed that TD min is defined by the potential planners depending on the location of the proposed rail transit system (i.e., urban or rural).

If the required minimum tunnel depth is not satisfied at any point along the line under consideration, even after replacing the maximum allowable value of G in equation (5.14) and (5.15) then the line is moved downwards by an amount that satisfies (TD min).

- **Step 5:** compute the tunnel depth of the next ranked line in the list generated in step 3 using the procedure described in step 4.
- **Step 6:** find the common stations between those lines whose tunnel depths have already been computed and rank them in a list based on their ground elevations in ascending order.
- **Step 7:** for each station in the common station list created in the previous step, check whether their respective lines are crossing each other or not, and perform the following "what if" analysis. Obviously, the algorithm starts with the station that has a zero index in the list (i.e., the first station in the list).
  i. If the corresponding lines of the common stations cross each other then they must be placed at different elevations so that their associated trips do not influence each another. This will eliminate the need of installing traffic signal controls in the case where the two crossing lines are at the same level (elevation). The required minimum elevation difference (ΔTE min) between any two crossing lines is assumed to be defined by the user/ planner depending on the ground conditions. The algorithm is therefore, designed to check whether (ΔTE min) between any two crossed lines is satisfied or not. If it is satisfied, then the algorithm proceeds to check the corresponding line of the next station in the common station list. Otherwise, the algorithm changes the value of G between G_{max} and G_{min} systematically to satisfy the required (ΔTE min) while ensuring that TD min remains

constant. In the case where changing the values of G does not work then one of the crossing lines is moved downwards by an amount that satisfies ΔTE min. It is very important to note that the moved line is the line that is currently under the consideration (i.e., the line that is taken at step 5). Also it is worth noting that in the cases where there are more than two lines crossing each other at the common station, the algorithm checks ΔTE min between any two crossed lines individually in the same manner, so as to ensure that ΔTE min is satisfied along all the crossed lines in the network.

ii. If the respective lines of the common station do not cross, the algorithm proceeds to check the corresponding line of the next station in the common station list.

- **Step 8:** repeat steps 5 through to 7 to compute the tunnel depth along all the lines of the proposed alternative solutions.

Once the tunnel depth has been computed, the construction cost of the tunnel is calculated. This is done by first using the calculated tunnel depth and the geological map of the potential study area to determine the ground conditions (i.e., soil and rock) through which the tunnels pass, and then applying the relevant expressions in table 5.2. The geological map, which provides preliminary information on the ground layers of the potential study area and associated properties and depth from the ground surface, are fed into the evaluation framework of the proposed integrated optimization algorithm via interaction with the GIS.

The tunnel cost of each line in the network is calculated by summing up the tunnel cost of the corresponding segments. Each segment is divided into two sub-segments, with one lying between the start station and the middle point of the segment and the other lying between the middle point and the end station of the segment. The ground condition of each sub-segment is determined at both its ends using the respective calculated tunnel depth and spatial geological information. If both ends have the same ground condition (i.e., soil or rock) then the respective tunnel cost is calculated using the associated cost estimate models presented in table 5.2. That is, when both ends are soil then the tunnel cost is calculated using the soil cost models, and when both ends are rock then the tunnel cost is calculated using the rock cost models. However, if the two ends are of different ground conditions, i.e., one is soil while the other is rock then the respective tunnel cost is calculated by averaging the construction cost obtained from the application of both soil and rock cost models. In order to specify exactly which of the presented cost models in table 5.2 is to be applied, it is also essential to determine the method of the tunnel excavation (i.e., conventional or mechanized). The determination of the tunnel excavation method depends on several factors, such as geological conditions and geographical location, as well as the size, length, shape, orientation and end use/final application of the tunnel. In this study, however, for simplicity, it is assumed that the excavation method is selected by planners prior to the evaluation process.

5.4.3.3 Escalator Barrel Cost

Similar to the tunnel construction cost, the construction cost of the escalator barrel is calculated after computing the tunnel depth using the proposed expressions in table 5.2. Basically, the escalator barrels connect station concourse halls and line platforms, and thus its length depends on the difference of the corresponding station concourse hall and station platform elevations, as well as the inclination angle of the escalator. As discussed in section 4.3.2, it is assumed that the station concourse halls are built at the ground level and thus the spatially respective ground elevation, which is extracted from the topography map of the study area via interaction with GIS, is used as its elevation. As for the station platform elevation, this is computed by subtracting the spatially respective ground elevation from the tunnel depth computed in section 5.4.3.2. Thus the escalator barrel length of each station in the proposed alternative solution (EL) is calculated as:

$$EL_{sti} = (GE_{sti} - TD_{sti}) \times \cos\theta \qquad (5.16)$$

Where:

GE_{sti} is the ground elevation of station i at the corresponding concourse hall,

TD_{sti} is the tunnel depth at station i,

$\cos\theta$ is the escalator inclination angle, which is assumed to be predefined by planners using the internationally standard values.

Once the escalator barrel length of each station has been calculated and the escalator barrel diameter defined, the expressions in table 5.2 are applied to compute the escalator barrel cost. The ground condition through which the escalator barrel is excavated is determined by comparing the spatially respective station's concourse hall and platform levels with the geological data. Similar to the tunnel construction cost, when both ends of the escalator barrel are soil then the soil cost models are used to estimate the escalator barrel cost, and the rock cost models are used only when both ends lie in rock. If, however, the two ends are of different ground conditions, i.e., one is soil while the other is rock, then the respective escalator barrel cost is calculated by averaging the cost obtained from the application of both soil and rock cost models.

5.4.3.4 Station Building and Equipping Cost

The cost of building and equipping stations with required facilities, which is classified as a size-dependent cost in this thesis, is a function of the size and the number of stations along the line network of the proposed rail system. Apart from the intersected stations, the size of all the stations are, however, assumed to be identical since all of them are planned to operate and serve the potential users of the system in a similar way. This is not to suggest that the intersected stations are planned to operate and offer services differently from the other stations in the system, but due to serving more than one line of the network the potential usage of these stations is often greater than the potential usage of the other stations. To accommodate such additional usage, and associated service facilities, their sizes are greater than the other stations, and accordingly the sizes of the intersected stations are assumed to be proportional to the number of lines passing through them. Accordingly, the cost of building and equipping stations is expressed as:

$$S_b = \sum_{l=1}^{q} (Ns \times C_{S_b}) \tag{5.17}$$

Where:

S_b is the total cost of building and equipping the stations with necessary facilities (£),

Ns is the number of stations along line l, and

C_{S_b} is the unit cost of building and equipping a station with required facilities (£).

5.4.3.5 Track Cost

The cost of constructing the rail track covers the cost of laying the tracks and fastening rails. This is considered to be a linear function of the line network length, and is thus calculated by multiplying the total line network length by the unit cost of track length.

$$T_r = \sum_{l=1}^{q}(Len_l \times C_{T_r}) \tag{5.18}$$

Where:

T_r is the total cost of laying rail tracks and (£),

Len_l is the length of line *l (m)*, and

C_{T_r} is the unit cost of laying rail tracks and fastening rails (£/m).

5.4.4 Rail Transit System Demand Forecast

The degree to which a rail transit system contributes to increasing transit usage and reducing reliance on the car is one of the most important criteria in planning such a system. This thesis employs a discrete mode choice model, which is a widely-accepted passenger forecasting model, to predict the user share of the proposed rail transit system and to identify how the proposed new service will influence the mode choice of existing travellers.

The discrete choice models of mode choice are used to analyse and predict a decision-maker's choice of one among a set of different discrete alternatives. These models are based on random utility theory, which assumes that a decision-maker's preference towards each alternative depends on the degree of satisfaction or utility measure associated with each alternative; i.e. that the choice is the alternative that maximizes the decision-maker's utility or satisfaction. The utility measure associated with alternatives is typically specified as a function of the alternatives' attributes and decision-makers' characteristics that describe their valuation for each alternative. However, many of the attributes and characteristics that influence decision-maker's utilities such as comfort and convenience cannot be observed or measured directly. In order to recognize and accommodate the effect of such unobserved attributes and characteristics on the internal decision-making process, the utility of the alternatives is defined as a random variable consisting of two components. One of the components of the utility function, which is called the deterministic or systematic portion of the utility, represents the observable and measurable attributes of alternatives and characteristics of decision makers. The other component, which is often called the stochastic error term of the utility function, represents the unknown and/ or unobservable attributes of alternatives and characteristics of decision makers. This can be expressed as:

$$U_{md} = V_{md} + \varepsilon_{md} \tag{5.19}$$

Where:

U_{md} is utility of the alternative m to the decision maker d,

V_{md} is the deterministic or observable utility of the alternative m to the decision maker d, and

ε_{md} is the error term or unobservable utility of the alternative m to the decision maker d.

Based on the data type, there are two basic approaches for choice modelling; aggregate and disaggregate. The disaggregate approach models the choice of individuals as a function of the attributes of the alternatives, and the socio-demographic characteristics of each individual among the individual's choice set (Koppelman and Bhat, 2006). The aggregate approach models the choice of all, or a selection of decision makers, as a function of attributes of the alternatives and the socio-demographic characteristics of the group among a set of choices. The aggregate choice modelling approach is employed here, since this thesis assumes that the trip matrix data is known from an external regional demand forecasting model, which provides the aggregate trips of all individuals for each traffic analysis zone. The proposed algorithm employs a logit choice model, the most widely used discrete choice model in real world practice, to forecast the user share of the proposed rail transit system. The structure and formulations of this model are presented in the next two subsections. It should be noted that since this model uses aggregate data, the term logit choice model in this thesis refers to the aggregate logit choice model.

5.4.4.1 The Logit Choice Model Structure

The logit model formula is derived from equation (5.19), assuming that the error terms are Gumbel distributed, and identically and independently distributed across both alternatives and individuals.

It is the most widely used discrete mode choice model because of its amenability to mathematical manipulation and closed form, which makes it easy to use and to interpret results, compared to other models. The model gives the probability of choosing each alternative as a function of its deterministic portion of the utility of all the alternatives. It has three important properties (Koppelman and Bhat, 2006): (1) the sigmoid or S shape relation of choice probabilities to representative utility. This implies that a small increase in the representative utility of an alternative will not significantly affect its probability of being chosen if the utility of this alternative is very low or high compared with other alternatives. The point at which a small improvement /increase in representative utility can tip the balance in people's choice and make a substantial change in the choice probability of the alternative is when the probability is 0.5, i.e., a 50–50 chance of the alternative being chosen; (2) the independence from irrelevant alternatives (IIA) property. That is, for any pair of alternatives, the ratio of the choice probabilities is unaffected by the availability and attributes of any other alternatives; and (3) the equivalent differences property. This states that the alternative choice probabilities rely only on the differences in the deterministic utilities of different alternatives, not on their numerical values. The premise is that an alternative is chosen if, and only if, it maximizes the individual's utility.

The available transport modes for travelling within the study area, the city of Leicester, are walk, bicycle, bus and car. The logit model, which is incorporated into the evaluation framework of the proposed algorithm, therefore predicts the market share for each mode including the new proposed rail system. Then the proposed algorithm uses these values to split the trip matrix of each origin destination pairs of TAZs in the study area into four different matrices, one for each mode, active (i.e. walk and bicycle), bus, car, and rail. Thereafter, it uses these trip matrices to compute both the passenger and operator costs discussed in section 5.4.1 and 5.4.2. Figure 5.9 illustrates the choice structure used in the logit model.

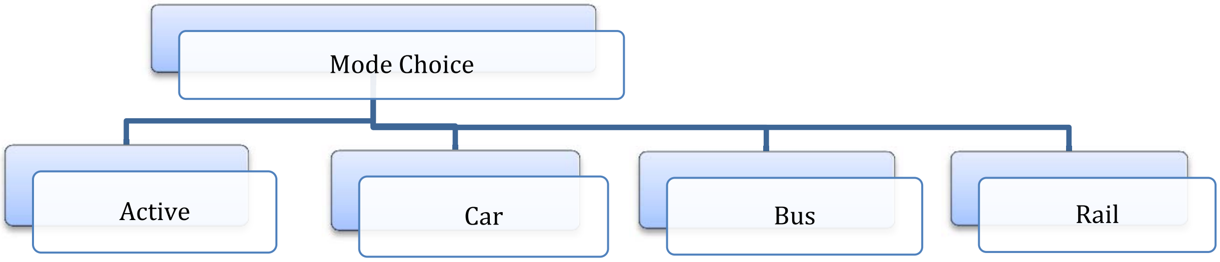


Figure 5.9: The Logit Model Structure

The probability of choosing mode m among different transport modes between any OD pairs ij is expressed as:

$$P_{ijm} = \frac{e^{-\lambda GC_{ijm}}}{\sum_m e^{-\lambda GC_{ijm}}} \tag{5.20}$$

Where: GC_{ijm} is the generalized cost of mode m for travelling from TAZ i to TAZ j, and λ is the sensitivity parameter to the generalized cost to be estimated. The next subsection details the estimation of this parameter (λ).

The deterministic utility is usually specified to be linear in parameters: $GC_{ijm} = \beta x_{ijm}$, where x_{ijm} is a vector of observed variables relating to alternative m, and the β are the associated parameters. With this specification, the probability of choosing mode m becomes:

$$P_{ijm} = \frac{e^{\beta x_{ijm}}}{\sum_m e^{\beta x_{ijm}}} \tag{5.21}$$

The vector of the observed variable is a mathematical function of the attributes of the alternatives and the characteristics of travellers that are measurable and are expected to affect travellers' preferences among alternatives. The alternatives' attributes include measures of in-vehicle travel time, access time, egress time, waiting time, search time for parking spaces and travel fares, while the characteristics of travellers include age, gender, income, car ownership and employment status. The case study in this thesis employs those variables that are incorporated into the choice model structure of the Leicester and Leicestershire Integrated Transport Model (LLITM) for describing the deterministic utility of the alternatives. The goals for LLITM are wide-ranging; mainly it is intended to evaluate interactions between land use and transport and predict the demand for travel throughout Leicestershire, as well as assessing transport users response to specific developments in the transport network (PR103 Demand Model Development Report, 2013). This model incorporates sets of variables that are related to the attributes of alternatives and the characteristics of travellers in its choice model framework. The representative variables of the attributes of alternatives that are considered for inclusion in the model involve in-vehicle travel time, access time, egress time, waiting time, search time for parking spaces, travel fares, fuel and other operating and maintenance costs. As for incorporating the variables associated with travellers' characteristics, the model groups travellers into nineteen segments according to their trip purposes and income levels, assuming that each of these segments are entirely separate and do not interact with one another. The segmentation categories are derived from nine trip purposes crossed by three income level categories. The travellers' trip purposes are grouped into nine segments as follow:

- commuting,
- education,
- shopping,
- home based employer business,
- home based other,
- non-home based employer business,
- non-home based other ,
- light goods vehicles (LGV), and
- other goods vehicles (OGV)

The travellers' income categories are grouped into three income segments:

- low income,
- medium income, and
- high income

In addition to this segmentation at the mode choice stage, the LLITM makes a distinction between travellers who have a car available for a trip (full-car) and those who have a car available with competition (part-car) or no car at all (no-car) and are thus restricted in their choice of modes.

Interaction of these segmentation variables with the service attributes offered by the various alternatives quantifies the systematic variability in travel choice decision.

The proposed optimization algorithm is designed to accommodate all the segments identified above into its mode choice model framework as well as including the variable set describing the attributes of the alternatives. However, it considers only home based trips, (i.e., commuting, education, shopping, home based employer business and home based other segments) and the medium income level segments to reduce the model running time. In addition, incorporating these segments into the mode choice modelling process should be sufficient to examine the effectiveness of the proposed model. Accordingly, for each segment considered for inclusion in the modelling of travellers' choices, the service attributes of the alternative modes are formulated in terms of generalized cost, which is a weighted linear combination of travel time and other costs of travel. It is essential that the time and monetary elements are measured in consistent units, either time or cost, so that the demand can be estimated directly to fall or rise with an increase or a reduction in either. In this thesis, the generalized cost of the alternative transport modes is measured in units of money (pounds sterling); that is the travel time elements associated with the alternative modes are converted to equivalent costs by multiplying them to values of time. The generalized cost formulations for the alternative modes are discussed below.

1- Rail Generalized Cost (V_{RailGC}): is comprised of access/egress time to/from rail stations, waiting time at stations, on train travel time and rail fare costs.

$$V_{RailGC} = \beta_1 x_{Rat} VOT + B_2 x_{Ret} VOT + \beta_3 x_{Rwt} VOT + x_{Rtt} VOT + x_{Rfare} \quad (5.22)$$

Where: x_{Rat} is access time to rail stations in minutes, x_{Ret} is egress time from rail stations in minutes, x_{Rtt} is on train travel time in minutes, x_{Rwt}, is waiting time at rail station in minutes, x_{Rfare} is rail fare in pence, VOT is the value of time, and β_{1-3} are weights of access, egress, and waiting time respectively.

The details on how the proposed model calculates these travel time elements is elaborated in section 5.4.1. The rail fare is calculated using the following formula, which is a linear function of travel distance. This function was developed based on the Central Leicestershire Transport Model's (CLTM) rail fare functions.

$$x_{Rfare} = A + B \times D_{Rtr} \quad (5.23)$$

Where: A is a fixed cost term, B is the cost of travel per kilometre, and D_{Rtr} is the rail travel distance in kilometres. Table 5.3 presents recommended values of the rail fare function parameters (i.e., A and B), which vary with travellers' trip purposes.

Table 5.3: The Rail Function Parameters (2001 prices)

Trip Purpose	A (pence)	B (pence/km)
Commuting	16.10	7.78
Business	19.17	9.26
Education	12.84	6.20
Other	16.29	7.87

Source: (The Regional PTOLEMY Model of the East Midlands Region Model Development and Validation Report, 2010)

As depicted in table 5.3, the value of both parameters are calibrated based on the 2001 prices and therefore the computed values of the rail fare will be based on the same year's price. This study therefore converts the computed rail fare values to 2015 prices using the GDP deflator recommended by (WebTAG Data Book, 2015) and then uses them for calculating the generalized cost of rail.

2- Bus Generalized Cost (V_{BusGC}): similar to the generic generalized cost of the rail trips, the bus generalized cost consists of access/egress time to/from bus stops, waiting time at bus stops, on bus travel time and bus fare costs.

$$V_{BusGC} = \beta_1 x_{Bat} VOT + B_2 x_{Bet} VOT + \beta_3 x_{Bwt} VOT + x_{Btt} VOT + x_{Bfare} \quad (5.24)$$

Where: x_{Bat} is access time to the bus stop in minutes x_{Bet} is egress time from the bus stop in minutes, x_{Bwt} is waiting time at the bus stop, x_{Btt} is on bus travel time in minutes, and x_{Bfare} is bus fare in pence. It is, however, assumed that the bus network associated travel time and fares of each TAZ pairs at different time periods are obtainable from external travel forecast models.

Car Generalized Cost (V_{CarGC}): is comprised of in-car travel time, time spent searching for a parking space and other car costs, which consists of three main components; toll and charge costs, fuel costs and non-fuel costs. The car fuel cost is a function of car travel speed, travel distance and fuel price, while the car non-fuel cost includes car maintenance and depreciation costs. The formulas for calculating these two car cost components can be found in the WebTAG data book- section A1.3.14 and A1.3.15 (WebTAG Data Book, 2015). Similar to the bus generalized cost, it is assumed that the road network travel time and other car cost information for each TAZ pair at different time periods are obtainable from external travel forecast models.

$$V_{CarGC} = \beta_4 x_{Cpt} VOT + x_{Ctt} VOT + x_{Ctc} \tag{5.25}$$

3- Walk and Bicycle Generalized Cost ($V_{ActiveGC}$): unlike other alternative modes, this consists of only the walking or bicycle riding time costs.

$$V_{ActiveGC} = \beta_5 x_{Awt} VOT \tag{5.26}$$

Where: x_{Awt} is walking or bicycle riding time in minutes and β_5 is the associated weight coefficient. Similar to both the car and bus generalized cost information, it is considered that the generalized cost information of this transport mode are also obtainable from external travel forecast models.

It is worth noting that this concept of the generalized cost of travel and its formulations with respect to different transport alternative modes are recommended by the UK Transport Analysis Guidance WebTAG (WebTAG: TAG Unit M2 Variable Demand Modelling, 2014) and is, therefore, used to build most transport demand models in the UK, including the LLITM. Consequently, the values of time (VOT) and the associated weights of the travel time elements (β_{1-5}) used in the proposed model were constrained to the recommended WebTAG values. Table 5.4 presents the values of time (VOT) recommended by WebTAG (unit 3.10.12), and used in LLITM. In this thesis these values were converted to 2015 prices using the GDP deflator recommended by (WebTAG Data Book, 2015) and used to calculate the generalized cost of the alternative modes. Values of out of vehicle time, which include access, egress, waiting and walking times, are correlated to the values of in-vehicle time by applying appropriate weights (β). The recommended values of these time elements according to WebTAG is 2, which implies that these time elements are valued at double the in-vehicle time.

Table 5.4: Values of Time (2010 prices)

Trip Purposes	Value of Time in Pence per Minute		
	Low Income	Medium Income	High Income
Commuting	8.619	11.081	14.074
Education	8.776	9.794	10.873
Shopping	8.776	9.794	10.873
Business	46.721	46.721	46.721
Other	8.776	9.794	10.873

Source: (PR103 Demand Model Development Report, 2013)

With the above generalized cost associated with the alternative modes, the probability of choosing mode m among the alternative modes between any OD pairs ij is computed applying equation (5.20). It is also very important to note that the formulation of the mode choice model is not tailored for the city of Leicester. The proposed optimization algorithm is designed in a way that it allows planners to define the related parameters of the alternatives' attributes and travellers' characteristics, which are proposed to the service area/ city by relevant transport entities.

5.4.4.2 Estimation of the Mode Choice Model Sensitivity Parameter

Using the dataset obtained from Leicester City Council, which was extracted from LLITM, this section estimates the values of lambda (λ). The data set consisted of:

1- Generalized cost matrices of each TAZ pairs (GC_{ijm}) at four different time periods (AM, OP, IP, PM) of a day in 2015 for each alternative transport mode in the city, i.e., car, bus and active mode.
2- Demand matrices of each TAZ pairs (P_{ijm}) at four different time periods (AM, OP, IP, PM) of a day in 2015 for each alternative transport mode in the city, i.e., car, bus and active mode.

Since the corresponding equation for lambda (equation 5.20) is in a nonlinear from, it is converted to a linear form, as below, to enable easier estimation of the linear regression:

Using equation (5.20) the proportion of travellers choosing each alternative is:

$$P_{Car} = \frac{e^{(-\lambda \times GC_{Car})}}{e^{(-\lambda \times GC_{Car})} + e^{(-\lambda \times GC_{Bus})} + e^{(-\lambda \times GC_{Active})}} \tag{5.27}$$

$$P_{Bus} = \frac{e^{(-λ \times GC_{Bus})}}{e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}} \quad (5.28)$$

$$P_{Active} = \frac{e^{(-λ \times GC_{Active})}}{e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}} \quad (5.29)$$

The proportion or probability of choosing each of the alternatives is calculated using the demand matrix dataset. This is by summing up the associated demand of the alternative modes (i.e., car, bus and walk and bicycle) first to obtain the total demand. The proportion of each alternative mode of each TAZ pair is then computed by dividing its corresponding demand by the calculated corresponding total demand. Now, taking the ratio of the above three mode proportions yields:

$$\frac{P_{Car}}{P_{Bus}} = \frac{e^{(-λ \times GC_{Car})}/[e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}]}{e^{(-λ \times GC_{Bus})}/[e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}]} = \frac{e^{(-λ \times GC_{Car})}}{e^{(-λ \times GC_{Bus})}} \quad (5.30)$$

$$\frac{P_{Car}}{P_{Active}} = \frac{e^{(-λ \times GC_{Car})}/[e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}]}{e^{(-λ \times GC_{Active})}/[e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}]} = \frac{e^{(-λ \times GC_{Car})}}{e^{(-λ \times GC_{Active})}} \quad (5.31)$$

$$\frac{P_{Bus}}{P_{Active}} = \frac{e^{(-λ \times GC_{Bus})}/[e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}]}{e^{(-λ \times GC_{Active})}/[e^{(-λ \times GC_{Car})} + e^{(-λ \times GC_{Bus})} + e^{(-λ \times GC_{Active})}]} = \frac{e^{(-λ \times GC_{Bus})}}{e^{(-λ \times GC_{Active})}} \quad (5.32)$$

Taking the logarithm of both sides of the above three equations and rearranging the yields:

$$Log\,\frac{P_{Car}}{P_{Bus}} = \frac{Loge^{(-λ \times GC_{Car})}}{Loge^{(-λ \times GC_{Bus})}} = λ(GC_{Bus} - GC_{Car}) \quad (5.33)$$

$$Log\,\frac{P_{Car}}{P_{Active}} = \frac{Loge^{(-λ \times GC_{Car})}}{Loge^{(-λ \times V_{ActiveGC})}} = λ(GC_{Active} - GC_{Car}) \quad (5.34)$$

$$Log\,\frac{P_{Bus}}{P_{Active}} = \frac{Loge^{(-λ \times GC_{Bus})}}{Loge^{(-λ \times GC_{Active})}} = λ(GC_{Active} - GC_{Bus}) \quad (5.35)$$

The form of the (5.33) to (5.35) equations are linear, and thus the linear regression method can now be applied to these equations for solving the value of lambda. Theoretically, linear regression assumes that the relationship between response and explanatory variables is linear, which means that the scatterplot between these variables closely resembles a straight-line. It is therefore, necessary to examine the nonlinearity and heteroscedasticity of the data before running the regression in order to ensure that the linear regression model fits the data well. Accordingly, residual plots are generated to identify patterns which may indicate problems of nonlinearity and heteroscedasticity. The residual plots of the data show discernible patterns; that is the residuals had no symmetrical patterns and a consistent spread throughout the centrelines. To improve the fitness of the regression models, the corresponding points of the extreme residuals were excluded for the regression analysis, and subsequently the R-square values improved substantially. The R-square values were 0.36, 0.15, 0.21, 0.32, and 0.14 for commuting, education, shopping, business and other home based trips, respectively, before omitting the outliers, whereupon they improved by more than 50% in all cases, as depicted in table 5.5.

In addition, it is important to note that the sensitivity of the logit model is likely to be different with different travel purpose (TAG UNIT M2- Variable Demand Modelling). The numerical value of the sensitivity parameter (lambda) is likely to be larger where there is more freedom to choose. For instance, more optional travel, such as shopping trips, tend to have a larger lambda value compared to less optional travel, such as business trips. Lambda is therefore estimated for each travel purpose segment, as illustrated in table 5.5.

Table 5.5: Estimated Values of Lambda

Trip Purposes	Value of Lambda	R^2
Commuting	0.22	0.7552
Education	0.173	0.5914
Shopping	0.186	0.6147
Business	0.05	0.7604
Other	0.161	0.5822

As expected, the results show that, compared to the other trip purposes, the shopping trips have the highest value of lambda. This is a self-explanatory result since as discussed above, the more freedom travellers have in deciding when and where to travel, the more elasticity they would have to choose and thus the higher the sensitivity of their choices to the generalized costs. In addition, the results of the regression models show relatively high values for the coefficient of determinations, R-squares, ranging between 0.6-0.75. This implies that about 60% - 75% of the variability of the response variable- the ratio of the choice probabilities between any pair of alternatives- can be explained by the variability of the explanatory variables; the differences in their corresponding generalized costs.

5.5 Evolution of the Alternative Solutions

As illustrated in figure 5.1, once a population of alternative solutions is initialized and the corresponding fitness is evaluated, the GA starts the evolution to direct the population towards the optimum solution. Typically, GAs implement the evolution by applying three genetic operators; selection, crossover and mutation. Obviously, these evolution operators control the search process and GA performance to efficiently converge towards a global optimum solution. In general, these operators, in particular the crossover and mutation operators are problem specific. Therefore, they must be designed to fit in with the problem under investigation in order to facilitate the efficiency of the search process and prevent convergence toward a local optimum solution. Consequently, this chapter proposes three genetic operators that fit in with the framework of the solution of the rail transit system planning. Figure 5.10 illustrates the process flow of these operators followed by three subsequent sections that explain their roles in directing the search process and detail how they are designed to explore the search space efficiently for the optimum solution.

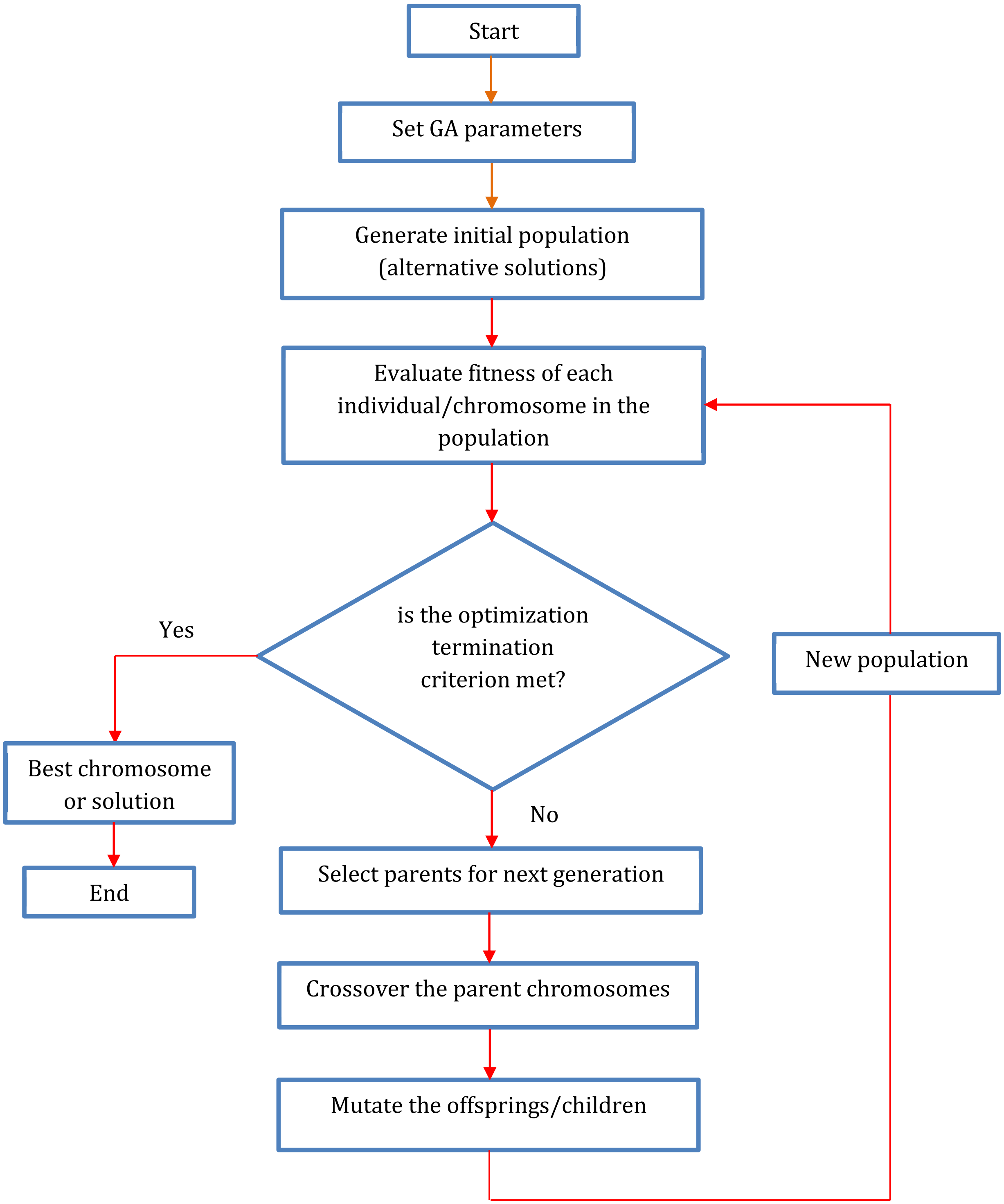


Figure 5.10: Flowchart of the Genetic Algorithm Process

5.5.1 Selection Operator

Basically, the selection operator determines which individuals are to be selected to form a mating pool for the next generation. The primary objective of this operator is to improve the average fitness of a population by giving individuals of higher fitness a higher chance to be selected for the next generation. Consequently, the selection operator directs the search to explore and exploit promising areas in the search space. The most widely used selection methods are roulette wheel selection, tournament selection and rank selection. Although each of these selection methods calculates the selection probability of individuals differently, they are all based on the same principle, which is survival for the fittest (i.e. fitter individuals have a higher chance of being selected for the next generation than weaker ones). Further details of these selection methods are available in many references (Blickle and Thiele, 1995, Goldberg and Deb, 1991). Obviously, different selection methods influence the performance of GA differently. Thus, close attention should be paid to choosing a selection method that ensures that the algorithm performs well, which is usually evaluated in terms of convergence rate and the required number of generations to arrive at the optimal solution. However, in this thesis, since the fitness values of the solutions can be very close, the tournament selection method is applied. This is because it is sensitive enough to differentiate the solutions of very similar fitness values. Tournament selection picks a subset of two solutions from the population randomly and then selects the best solution from this subset.

5.5.2 Crossover Operator

The crossover operator merges the individuals picked up by the selection operator to produce new individuals. Crossover operates on two individuals (parents) at a time and generates two offspring/children by exchanging the genetic information of both individuals (parents). The idea behind crossover is that the fitness of the generated offspring/children may be better than both of the parents if they take the best characteristics from each of the parents, thereby seeking to improve fitness and pulling the population as a whole towards the best possible solution. Crossover operation occurs at a probability rate denoted as Pc, which is defined as the probability of the number of generated offspring in each generation to the population size. This means that not all the selected individuals undergo the crossover operation unless the Pc value is equal to one. The value of Pc controls the expected number (Pc × Population size) of the generated offspring in each generation. A very high Pc value tends to explore more of the search space and mitigates the risk of trapping in local optima. On the other hand, it may result in a considerable wastage of computation time in exploring unnecessary areas of the search space. Therefore, it is very important to utilize a Pc value which maximizes the probability that the algorithm finding good solutions while converging speedily. The optimal value of Pc is problem specific(Patil and Pawar, 2015), and therefore, there is no fixed optimal value for Pc that can be generalized for the most real world problems. Accordingly, in the next chapter, this thesis examines different Pc values to determine the optimal value for solving the rail transit system planning problem.

In addition to the value of Pc, the crossover method greatly influences the performance of GA. To ensure the efficiency of the crossover operation, it is important to consider not only the value of Pc but also the method of the crossover. The most widely used crossover methods include one-point crossover, two-point-crossover and uniform-crossover. The complex configurations of this population's chromosomes and the satisfaction of the constraint set identified in section 3.4.2.2 during the evolution process, makes it difficult for these methods to produce good solutions. The configuration of the population's chromosomes is complex especially because the corresponding gene lengths, which represent the line network of the proposed rail system, are different from one another.

The length of each line in the network varies depending on the predefined minimum and maximum number of stations along the line and the predefined minimum and maximum distance in between the associated stations. Moreover, the bits order along the chromosome's genes, which represent the associated station locations of the line network, are dependent on each other. Therefore, in this thesis five different crossover operators are designed by adapting one-point and uniform crossover methods. These are then tested to determine which of them best fits the framework of rail transit system planning. The goodness of each of these operators is measured by its success rate (Csr), which is defined as the ratio of the total number of generated feasible offspring (Ctf) to the total number of parents undergoing the crossover (Ctm). Obviously, the crossover operation might produce infeasible offspring due to violation of the required constraints. Therefore, to handle these constraints the crossover operators are designed in such a way that they reject parental matings that result in infeasible offspring. The following subsections explain the design of these operators in detail and their implementation and goodness evaluations are presented in the next chapter.

5.5.2.1 Uniform Crossover at Chromosome Level (UCCL)

Adopting the conventional uniform crossover method, this operator is designed to merge parents at chromosome (line network) level to produce new offspring. The conventional uniform-crossover merges parent chromosomes by interchanging the genes of one parent to the corresponding genes of the other parent, with a probability of 0.5. In other words, the value of the second parent's gene is assigned to the first offspring and the value of the first parent's gene is assigned to the second offspring with a probability of 0.5. This operator applies the following steps for mating parents and producing new offspring:

- **Step 1:** Set two counters, one for counting the total number of matings (Ctm) that occur over the total number of generations, and the other for counting the feasible offspring (Ctf) produced from mating parents, i.e., set Ctm= 0 and Ctf=0.
- **Step 2:** Select two parents from the mating pool produced by the selection operator.
- **Step 3:** Generate a random number between 0 and 1 and compare it with the predefined value of Pc. If the generated number is greater than the value of Pc, go back to step 2 to select another two parents for mating. Otherwise, proceed with the following steps.
- **Step 4:** Determine gene bits to be mated along with one of the selected two parents. This is by successively visiting each gene bit of the parent and generating a random number between 0 and 1. If the generated random number is equal or smaller than 0.5, assign the gene bit to be mated and then proceed to the next gene bit. If the generated random number is greater than 0.5 proceed to the next gene bit. It is important to note that the proposed GA treats each alternative solution as a chromosome, the associated lines as genes, and the corresponding stations of the lines as gene bits, as mentioned earlier in section 5.3.
- **Step 5:** Simultaneously swap the gene bits identified for mating in step 4 with the corresponding gene bits of the other parent to produce two offspring, and increase the (Ctm) counter by one (i.e., Ctm= Ctm+1). As depicted in figure 5.11, the values of the second parent's

gene bits are assigned to the first offspring and the value of the first parent's gene bits are assigned to the second offspring.

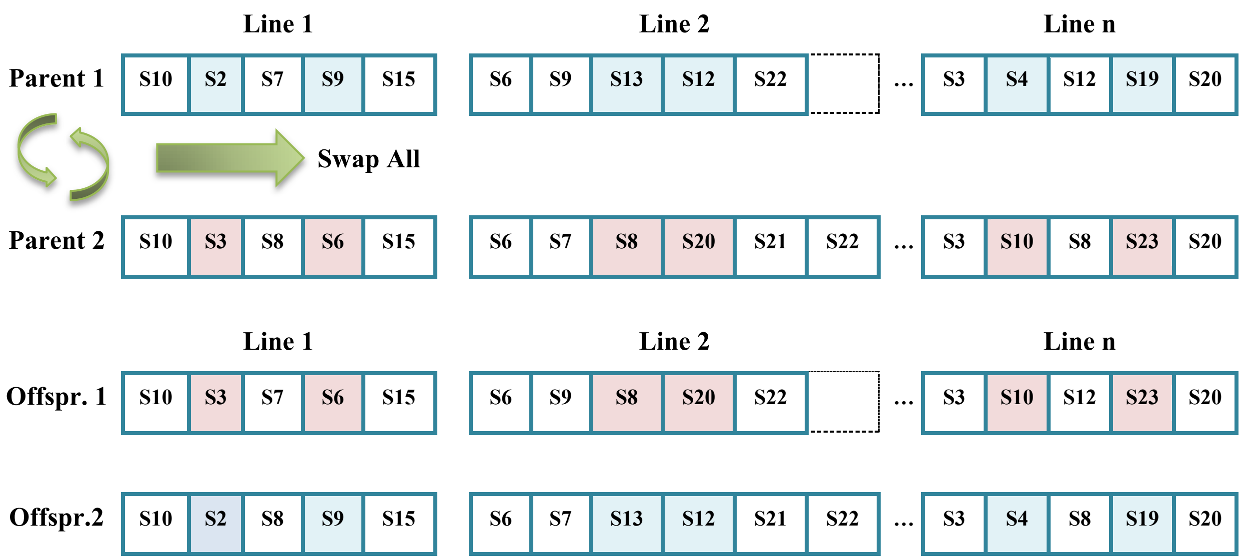


Figure 5.11: Illustration of the Uniform Crossover at Chromosome Level Operator

- **Step 6:** Check the feasibility of the generated offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario:
  - If the generated offspring satisfy the required constraints, increase the (Ctf) counter by one (i.e., Ctf = Ctf +1) and then repeat steps 2 to 6 to select the next two parents from the mating pool and proceed with the mating process.
  - Else (i.e., if the generated offspring violate the constraints) reject the mating and return the offspring chromosomes back to their initial state before the mating. Then repeat steps 4 to 6 to take another mating trail. If, after a predefined number of trails, the produced offspring are infeasible, return the offspring chromosomes back to their initial state before the mating, and then repeat steps 2 to 6 to select the next two parents from the mating pool and proceed with the mating process. The reason of performing these trails is that this operator simultaneously swaps all the selected gene bits of the parents, which makes the chance of producing offspring that satisfy all the required constraints very low. These trails are therefore, incorporated into this operator to allow efficient exploration of the search space

and avoid premature convergence by giving the parents more chance to produce new feasible offspring. The number of trails is set to be equal to the length of the parent chromosome, which is equal to the number of line networks connecting the stations. There is no specific reason for setting the number of trails to equal to this figure. However, the higher the number of trails the greater the chance of producing feasible offspring and thus the more thorough the exploration of the search space.

- **Step 7** repeat steps 2 to 6 until all the population's individuals picked up by the selection operator are mated.
- **Step 8:** Calculate the success rate (Csr) by dividing the value of Ctf by the value of Ctm. It should be noted that this value is calculated after completing the required number of generations.

5.5.2.2 Uniform Crossover at Gene Level (UCGL)

This operator is designed to merge parents at gene (individual line) level to produce new offspring. The main difference between this operator and the UCCL operator is in the combining level of the parent chromosomes. This operator combines the parents at gene level, which represents a single line in the proposed alternative solution, while the other operator combines the parents at chromosome level, which represents all the lines in the proposed alternative solution. The following key steps detail the procedure of this operation:

- **Step 1:** Set two counters, one for counting the total number of matings (Ctm) that occur over the total number of generations, and the other for counting the feasible offspring (Ctf) produced from the mating parents, i.e., set Ctm= 0 and Ctf=0.
- **Step 2:** Select two parents from the mating pool produced by the selection operator.
- **Step 3:** Generate a random number between 0 and 1, and compare it with the predefined value of Pc. If the generated random number is greater than the value of Pc, go back to step 2 to select another two parents for mating. Otherwise, proceed with the following steps.
- **Step 4:** Determine the gene bits to be mated along each gene (line) of one of the selected two parents, successively. This is by successively visiting the corresponding bits of the gene under consideration and generating a random number between 0 and 1. If the generated number is equal to or smaller than 0.5, assign the gene bit to be mated and then proceed to the next gene bit. If the generated number is greater than 0.5 proceed to the next gene bit.

- **Step 5:** Swap the gene bits identified for the mating in step 4 with the corresponding gene bits of the other parent to produce two offspring simultaneously, as illustrated in figure 5.12a , and increase the Ctm counter by one (i.e., Ctm= Ctm+1). As illustrated in figure 5.12b, the selected bit values of the second parent's gene are assigned to the first offspring and the selected bit values of the first parent's gene are assigned to the second offspring. Then check the feasibility of the generated offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario.
  - If the generated offspring satisfy the required constraints, increase the Ctf counter by one (i.e., Ctf = Ctf +1) and go back to step 4 to mate the next genes. It is very important to note that, apart from the mated genes, the other genes of the produced offspring are just a copy of their corresponding parent genes, as illustrated in figure 5.12b and 5.12c.
  - Else (i.e., if the generated offspring violate the constraints) reject the gene mating and return the mated genes back to their initial state before the gene mating. Then take another mating trail for the same parents' genes in the similar manner. If after a predefined number of trails the produced offspring are infeasible, return the genes back to their initial state before the gene mating and then go back to step 3 to mate the next genes (see figure 5.12b and 5.12c). The reason for performing a number of trails is to give the parents more chance of producing new feasible offspring and thus to ensure efficient exploration of the search space and avoid premature convergence. Similar to the UCCL operator, the number of trails is set to be equal to the length of the parent chromosome, which is obviously equal to the number of line networks connecting the stations.
- **Step 6:** repeat steps 4 and 5 until all corresponding genes of the selected two parents are mated and the final offspring are produced, as illustrated in figure 5.12d.
- **Step 7:** repeat steps 2 to 6 until all the population's individuals identified by the selection operator are mated.
- **Step 8:** Calculate the success rate (Csr) by dividing the value of Ctf by the value of Ctm. It should be noted that this value is calculated after completing the required number of generations.

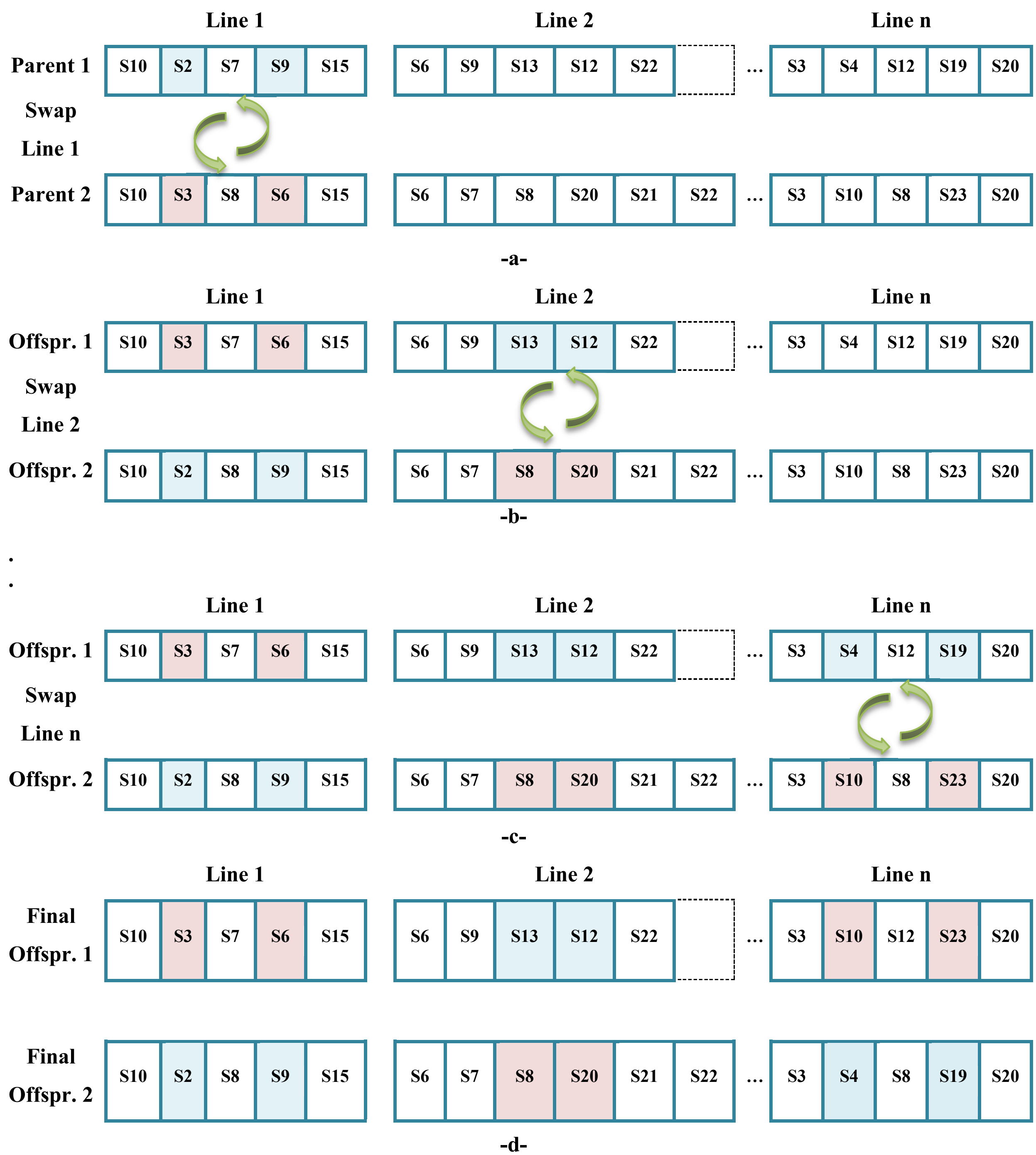


Figure 5.12: Illustration of the Uniform Crossover at Gene Level Operator

5.5.2.3 Uniform Crossover at Gene Bit Level (UCGBL)

This operator is designed to merge parents at gene bit (station) level to produce new offspring. The main difference between this operator and the two preceding operators, UCCL and UCGL, is again in the combining level of the parents. This operator combines the parents at gene bit (station) level while the other two operators combines the parents at chromosome (line network) and gene (single line) levels respectively. The following key steps describe the procedure of this operator:

- **Step 1:** Set two counters, one for counting the total number of matings (Ctm) that occur over the total number of generations, and the other for counting the feasible offspring (Ctf) produced from the mating parents, i.e., set Ctm= 0 and Ctf=0.
- **Step 2:** Select two parents from the mating pool produced by the selection operator.
- **Step 3:** Generate a random number between 0 and 1, and compare it with the predefined value of Pc. If the generated random number is greater than the value of Pc, go back to step 2 to select another two parents for mating. Otherwise, proceed with the following steps.
- **Step 4:** Successively determine the gene bits to be mated from one of the selected two parents. This is achieved by successively visiting the corresponding gene bits of the parent and generating a random number between 0 and 1. If the generated number is greater than 0.5 proceed to the next gene bit. If the generated number is equal or smaller than 0.5, assign the gene bit to be mated and then proceed with the following steps.
- **Step 5:** swap the gene bit identified for mating in step 4 with the corresponding gene bit of the other parent to produce two offspring, as illustrated in figure 5.13a, and increase the Ctm counter by one (i.e., Ctm= Ctm+1). Then check the feasibility of the generated offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario.
  - If the generated offspring satisfy the required constraints, increase the Ctf counter by one (i.e., Ctf = Ctf +1) and go back to step 4 to mate the next gene bits. It is very important to note that, apart from the mated gene bits, the other gene bits of the produced offspring are copies of their corresponding parental gene bits, as illustrated in figure 5.13b, 5.13c and 5.13d.
  - Else (i.e., if the generated offspring violate the constraints) reject the gene bits' mating and return the mated gene bits back to their initial state before the mating and then go back to step 4 to mate the next gene bits, as illustrated in 5.13b and 5.13c.
- **Step 6:** repeat steps 4 and 5 until all the corresponding gene bits of the selected two parents are mated and the final offspring are produced, as illustrated in figure 5.13d.

- **Step 7:** repeat steps 2 to 6 until all the population's individuals identified by the selection operator are mated.
- **Step 8:** Calculate the success rate (Csr) by dividing the value of Ctf by the value of Ctm. It should be noted that this value is calculated after completing the required number of generations.

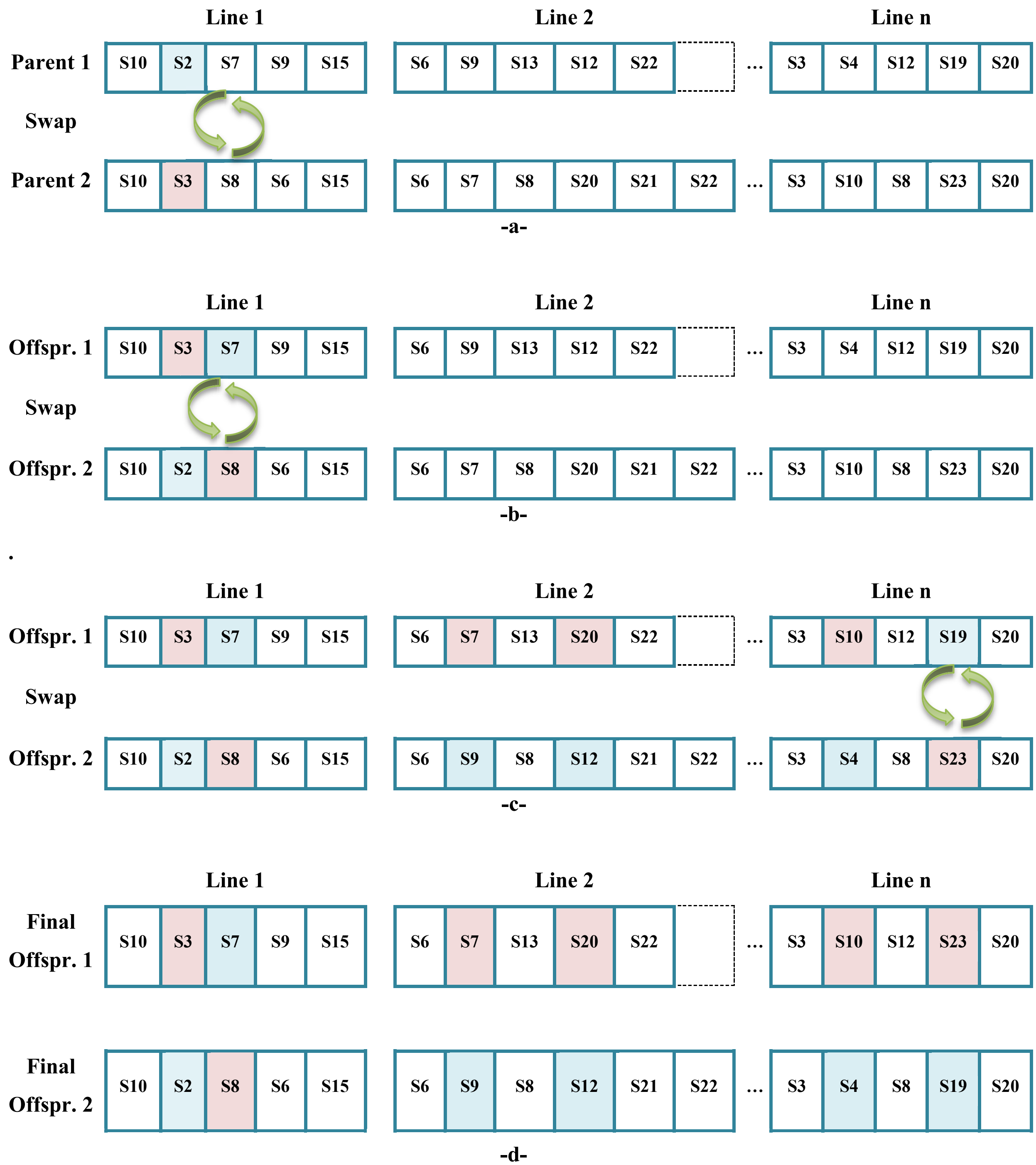


Figure 5.13: Illustration of the Uniform Crossover at Gene Bit Level Operator

5.5.2.4 One-Point Crossover at Chromosome Level (OPCCL)

This operator is designed to merge parents at chromosome (line network) level to produce new offspring. The only difference between this operator and the UCCL operator is in the employed crossover method for mating parents. The OPCCL operator adopted the conventional one-point crossover method, while the UCCL operator adopted a uniform crossover method. The one-point crossover randomly selects a common crossover point within two parents and then swaps the corresponding parts after the crossover point to produce two new offspring. Unlike one-point, uniform-crossover merges parent chromosomes at a gene bit level rather than gene segment level to produce new offspring chromosomes.

Adopting the one-point crossover method, this operator applies the following steps for mating parents and producing new offspring. However, it is important to note that, unlike the conventional one-point crossover, which swaps the corresponding parts of the parents after the crossover point, the form of the operator applied here swaps the corresponding parts of the parents *before* the crossover point. This is due to the complex configuration of the population's chromosomes which have different gene lengths from one another, as noted section 5.5.2 which may result in the production of infeasible offspring in most cases if the conventional one-point crossover method is employed. The following steps detail the procedure of this operator.

- **Step 1:** Set two counters, one for counting total number of matings (Ctm) that occur over the total number of generations, and the other for counting the feasible offspring (Ctf) produced from the mating parents, i.e., set Ctm= 0 and Ctf=0.
- **Step 2:** Select two parents from the mating pool produced by the selection operator.
- **Step 3:** Generate a random number between 0 and 1, and compare it with the predefined value of Pc. If the generated number is greater than the value of Pc, go back to step 2 to select another two parents for mating. Otherwise, proceed with the following steps.
- **Step 4:** Determine the gene bits to be mated along one of the two parents. This is achieved by visiting each gene of the selected parent successively and generating a random number between 0 and the corresponding length value of the particular gene under consideration. Denote the number generated when visiting each gene as Cp. As mentioned earlier, the proposed GA treats each individual line in the proposed alternative solution as a gene, and

therefore, the corresponding length is equal to the number of stations that are located along it. Then assign the corresponding gene bits of each gene whose index values are equal to or less than the corresponding Cp to be mated, as illustrated in figure 5.14.

- **Step 5:** Simultaneously swap the gene bits identified for mating in step 4 with the corresponding gene bits of the other parent to produce two offspring and increase the Ctm counter by one (i.e., Ctm= Ctm+1). As depicted in figure 5.14, the corresponding parts before the generated crossover points (Cps) are exchanged simultaneously to produce two new offspring.

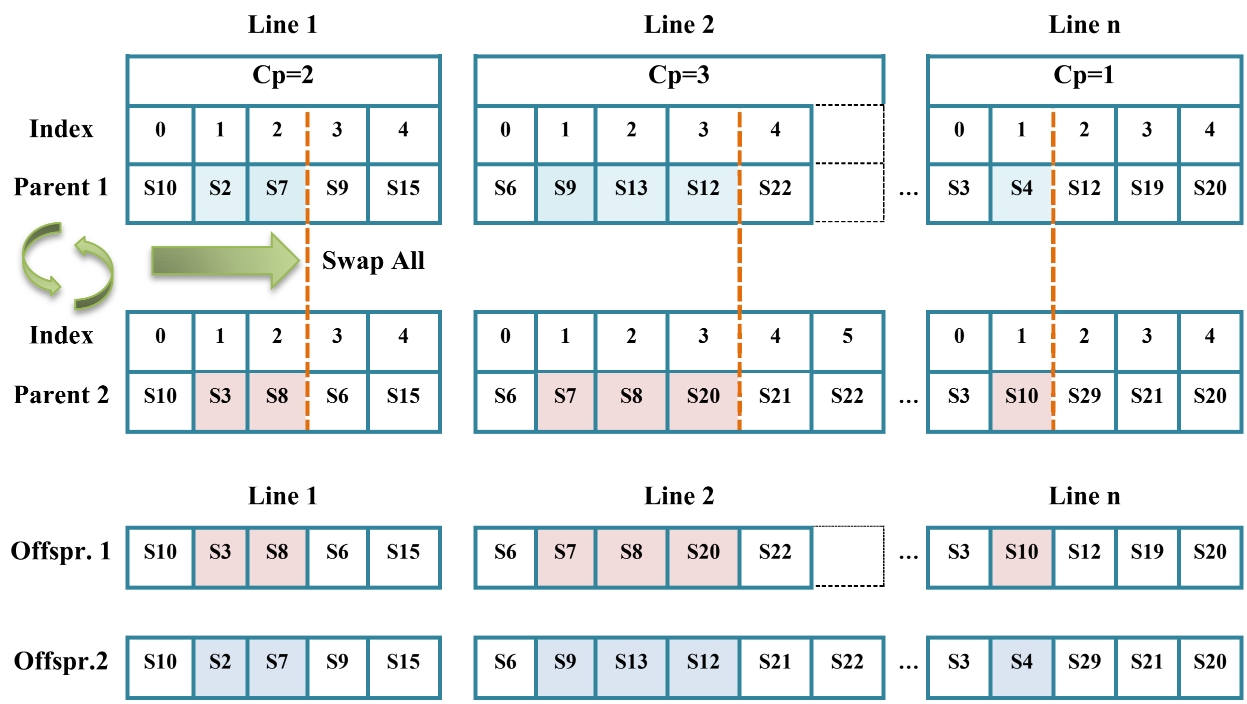


Figure 5.14: Illustration of the One-Point Crossover at Chromosome Level Operator

- **Step 6:** Check the feasibility of the generated offspring with respect to the required constraints, identified in section 3.4.2.2, and apply the following if-else scenario:
  - If the generated offspring satisfy the required constraints, increase the Ctf counter by one (i.e., Ctf = Ctf +1) and then go back to step 2 to mate the next parents.
  - Else (i.e., if the generated offspring violate the constraints) reject the mating and return the offspring chromosomes back to their initial state before the mating. Then repeat steps 4 to 6

to take another mating trail. If after a predefined number of trails the produced offspring are infeasible, return the offspring chromosomes back to their initial state before the mating, and then go back to step 2 to mate the next parents. Similar to the other preceding operators, the number of trails is set to be equal to the length of the parent chromosome.

- **Step 7:** repeat steps 2 to 6 until all the population's individuals identified by the selection operator are mated.
- **Step 8:** Calculate the success rate (Csr) by dividing the value of Ctf by the value of Ctm. It should be noted that this value is calculated after completing the required number of generations.

5.5.2.5 One-Point Crossover at Gene Level (OPCGL)

This operator is designed to merge parents at gene (individual line) level to produce new offspring. The main difference between this operator and the OPCCL operator is that it combines the parents at gene level (i.e., individual line level) while the other operator combines the parents at chromosome level (i.e., line network level). The following key steps detail this crossover:

- **Step 1:** Set two counters, one for counting total number of matings (Ctm) that occur over the total number of generations, and the other for counting the feasible offspring (Ctf) produced from the mating parents, i.e., set Ctm= 0 and Ctf=0.
- **Step 2:** Select two parents from the mating pool produced by the selection operator.
- **Step 3:** Generate a random number between 0 and 1, and compare it with the predefined value of Pc. If the generated random number is greater than the value of Pc, go back to step 2 to select another two parents for mating. Otherwise, proceed with the following steps.
- **Step 4:** Determine the gene bits to be mated along the corresponding genes of one of the two parents successively. This is achieved by visiting each gene of the selected parent successively and generating a random number between 0 and the corresponding length value of the particular gene under consideration. Denote the generated number for each gene as Cp. As noted earlier, the length of each gene is equal to the number of the corresponding gene bits (stations). Then assign the corresponding gene bits of the gene under consideration whose index value is equal to or less than Cp to be mated (see figure 5.15a) and proceed with the following steps.

- **Step 5:** Simultaneously swap the gene bits identified for mating in step 4 with the corresponding gene bits of the other parent to produce two offspring, and increase the Ctm counter by one (i.e., Ctm= Ctm+1). As depicted in figure 5.15b, the corresponding parts before the crossover point (Cp) of the gene under consideration are exchanged simultaneously while the other unmated genes remain unchanged at this step. Then check the feasibility of the generated offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario.
  - If the generated offspring satisfy the required constraints, increase the Ctf counter by one (i.e., Ctf = Ctf +1) and go back to step 4 to mate the next gene.
  - Else (i.e., if the generated offspring violate the constraints) reject the gene mating and return the mated genes back to their initial state before mating. Then take another mating trail for the same parents' genes in a similar manner. If after a predefined number of trails the produced offspring are infeasible, return the genes back to their initial state before the mating, and then go back to step 4 to mate the next gene of the selected parents (see figure 5.15b and 5.15c). Similar to the other operators, the number of trails is set to be equal to length of the parent chromosome.
- **Step 6:** repeat steps 4 and 5 until all corresponding genes of the selected two parents are mated and the final offspring are produced, as illustrated in figure 5.15d.
- **Step 7:** repeat steps 2 to 6 until all the population's individuals identified by the selection operator are mated.
- **Step 8:** Calculate the success rate (Csr) by dividing the value of Ctf by the value of Ctm. It should be noted that this value is calculated after completing the required number of generations.

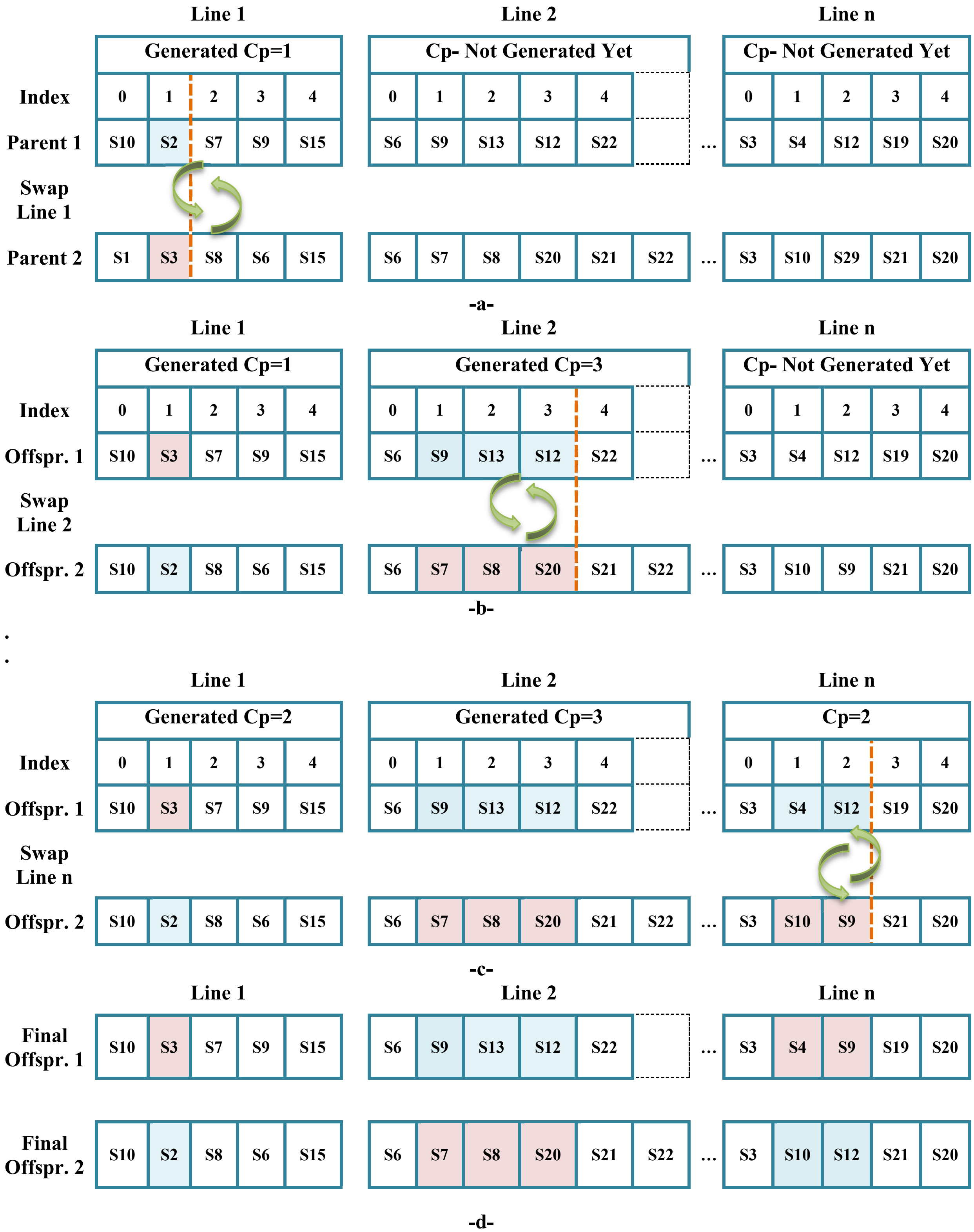


Figure 5.15: Illustration of the One-Point Crossover at Gene Level Operator

5.5.3 Mutation Operator

The mutation operator alters the value of one or more genes of the offspring pool generated by the crossover operator and alters a part of their content. It is used to produce new genetic information to ensure that the algorithm diversifies over the search space so as to avoid the premature convergence towards suboptimal solutions. The mutation operator works on a single individual/offspring at a time and produces a new individual/ offspring, and it occurs at a probability rate denoted as Pm. The mutation rate, or mutation probability, controls the frequency with which the mutation operator is applied. In general, a very high value of Pm tends to lead to a significant diversification of the search space. This may, in turn, pull good solutions away from the population and thus prevents the population from converging at any optimal solution. On the other hand, if the value of Pm is very low, the search mostly converges prematurely towards suboptimal solutions due to the higher exploitation rate. Therefore, it is very important to utilize a Pm value that maximizes the probability that the GA finds the global or near-global optimum solution of the problem. Since the optimal value of Pm varies for different problems, as (Patil and Pawar, 2015Patil and Pawar, 2015, Piszcz and Soule, 2006, Srinivas and Patnaik, 1994) have claimed, there is no consistent optimal value of Pm for most real world problems. Therefore, this thesis tests different values of Pm in the next chapter to find the optimal Pm value to solve the rail transit system planning problem.

In addition to the Pm value, the performance of GA depends to a great extent on the strategy or method of the mutation. Therefore, it is important to consider not only the value of Pm but also the method of the mutation while designing the GA operators. The design of a good mutation largely depends on the nature of the problem to be solved, which is the reason why this operator is often devised according to the complexity of the problem and the way that its solutions are represented. Consequently, this thesis designs three different mutation operators, and subsequently tests them to determine which of them best fits the framework of the rail transit system planning. Similar to the crossover operators, the goodness of each of these operators is measured by its success rate (Msr) which is defined as the ratio of the total number of the generated new feasible offspring (Mtf) to the total number of the mutated offspring (Mtm). As the mutation might produce infeasible offspring due to violation of the required constraints, the mutation operators are designed to reject the mutation operation if the constraints result in the production of infeasible offspring. The following subsections explain the design of these operators in detail. Their implementation and evaluation are addressed in the next chapter.

5.5.3.1 Mutation at Chromosome Level (MCL)

This operator is designed to take individuals from the offspring pool and randomly mutate them at the chromosome level (line network) to produce new offspring. It selects a number of gene bits within the individual with probability Pm, and simultaneously replaces them with randomly selected gene bits to produce a new offspring. The key steps of this mutation are detailed below:

- **Step 1**: Set two counters, one for counting the total number of mutations (Mtm) that occur over the total number of generations, and the other for counting the new feasible offspring (Mtf) produced from mutating the individuals within the offspring pool, i.e., set Mtm= 0 and Mtf=0.
- **Step 2:** Select an individual from the offspring pool produced by the crossover operator.
- **Step 3:** Determine the gene bits to be mutated along the selected offspring. This is achieved by successively visiting each gene bit of the offspring and generating a random number between 0 and 1. If the generated random number is greater than Pm, proceed to the next gene bit. Otherwise, assign the gene bit to be mutated and then proceed to the next gene bit.
- **Step 4:** replace the gene bits identified for the mutation in step 3 by randomly selecting gene bits to produce a new offspring , as illustrated in figure 5.16, and increase the Ctm counter by one (i.e., Ctm= Ctm+1).

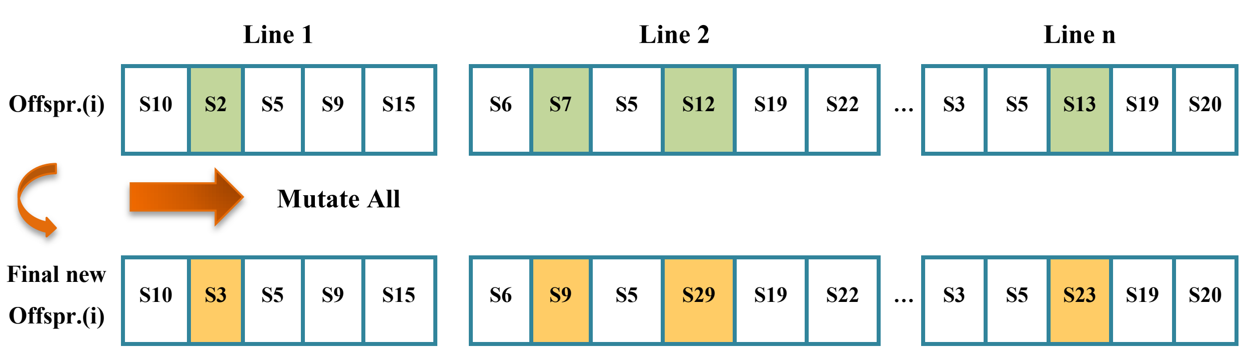


Figure 5.16: Illustration of Mutation at Chromosome Level Operator

- **Step 5:** Check the feasibility of the generated new offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario:
  - If the generated new offspring satisfy the required constraints, increase the Mtf counter by one (i.e., Mtf = Mtf +1) and then repeat steps 2 to 5 to select the next individual from the offspring pool and proceed with the mutation process.
  - Else (i.e., if the generated new offspring violate the constraints) reject the mutation and return the offspring chromosome back to its initial state before the mutation. Then repeat steps 3 to 5 to take another mutation trail. If after a predefined number of trails the produced offspring is infeasible, return the offspring chromosomes back to its initial state before the mutation, and then repeat steps 2 to 5 to select the next individual from the offspring pool and proceed with the mutation process. The reason for performing these trails is that this operator alters the selected gene bits for the mutation simultaneously which makes the chance of returning infeasible new offspring very high. These trails are therefore incorporated into this operator to give the individuals more chance to produce new feasible offspring so as to prevent the loss of diversity over the search space and avoid premature convergence. Similar to the crossover operators, the number of trails is set to be equal to length of the offspring, which is obviously equal to the number of line networks connecting the stations.
- **Step 6:** repeat steps 2 to 5 until all the population's individuals in the offspring pool are mutated.
- **Step 7:** Calculate the success rate (Msr) by dividing the value of Mtf by the value of Mtm. It should be noted that this value is calculated after completing the required number of generations.

5.5.3.2 Mutation at Gene Level (MGL)

This operator is designed to take individuals from the offspring pool and randomly mutate them at the gene level (individual line) to produce new offspring. It subsequently selects a number of gene bits within the individual's chromosome gene with probability Pm, and simultaneously replaces them by randomly selected gene bits to produce new offspring. The main difference between this operator and the MCL operator is that it mutates the individuals at the gene level, which represents a single line in the proposed alternative solution, while the other operator mutates the individuals at the chromosome level, which represents all the lines in the proposed alternative solution. The key steps of this mutation are detailed below:

- **Step 1**: Set two counters, one for counting the total number of mutations (Mtm) that occurs over the total number of generations, and the other for counting the new feasible offspring (Mtf) produced by mutating the individuals within the offspring pool, i.e., set Mtm= 0 and Mtf=0.
- **Step 2:** Select an individual from the offspring pool produced by the crossover operator.
- **Step 3:** Determine the gene bits to be mutated along each gene (individual line) of the selected offspring successively. This is achieved by visiting the corresponding bits of the gene under consideration successively and generating a random number between 0 and 1. If the generated random number is greater than Pm, proceed to the next gene bit. Otherwise, assign the gene bit to be mutated and then proceed with the next gene bit.
- **Step 4:** replace the gene bits identified for mutation in step 3 by randomly selected gene bits to produce a new offspring, as illustrated in figure 5.17a, and increase the Mtm counter by one (i.e., Mtm= Mtm+1). Then check the feasibility of the generated offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario
  - If the generated new offspring satisfy the required constraints, increase the Mtf counter by one (i.e., Mtf = Ctf +1) and go back to step 3 to mutate the next gene (see figure 5.17b). It is very important to note that, apart from the mutated genes, the other genes of the produced new offspring are copies of the corresponding individual genes, as illustrated in figure 5.17b and 5.17c.
  - Else (i.e., if the generated new offspring violate the constraints) reject the gene mutation and return the mutated gene back to its initial state before the mutation. Then take another mutation trail for the same individual's gene in a similar manner. If after a predefined number of trails the new offspring produced is infeasible, return the gene back to its initial state before the mutation, and then go back to step 3 to mutate the next genes (see figure 5.17b and 5.17c). The reason for performing a number of trails is to give the offspring more

chance of producing new feasible offspring so as to ensure efficient exploration of the search space and avoid premature convergence. Similar to the MCL operator, the number of trails is set to be equal to length of the parent chromosome, which is obviously equal to the number of line networks connecting the stations.

- **Step 5:** repeat steps 3 and 4 until all corresponding genes of the selected individual are mutated and the final new offspring (see figure 5.17d) is produced.
- **Step 6:** repeat steps 2 to 5 until all the selected population's individuals in the offspring pool are mutated.
- **Step 7:** Calculate the success rate (Msr) by dividing the value of Mtf by the value of Mtm. It should be noted that this value is calculated after completing the required number of generations.

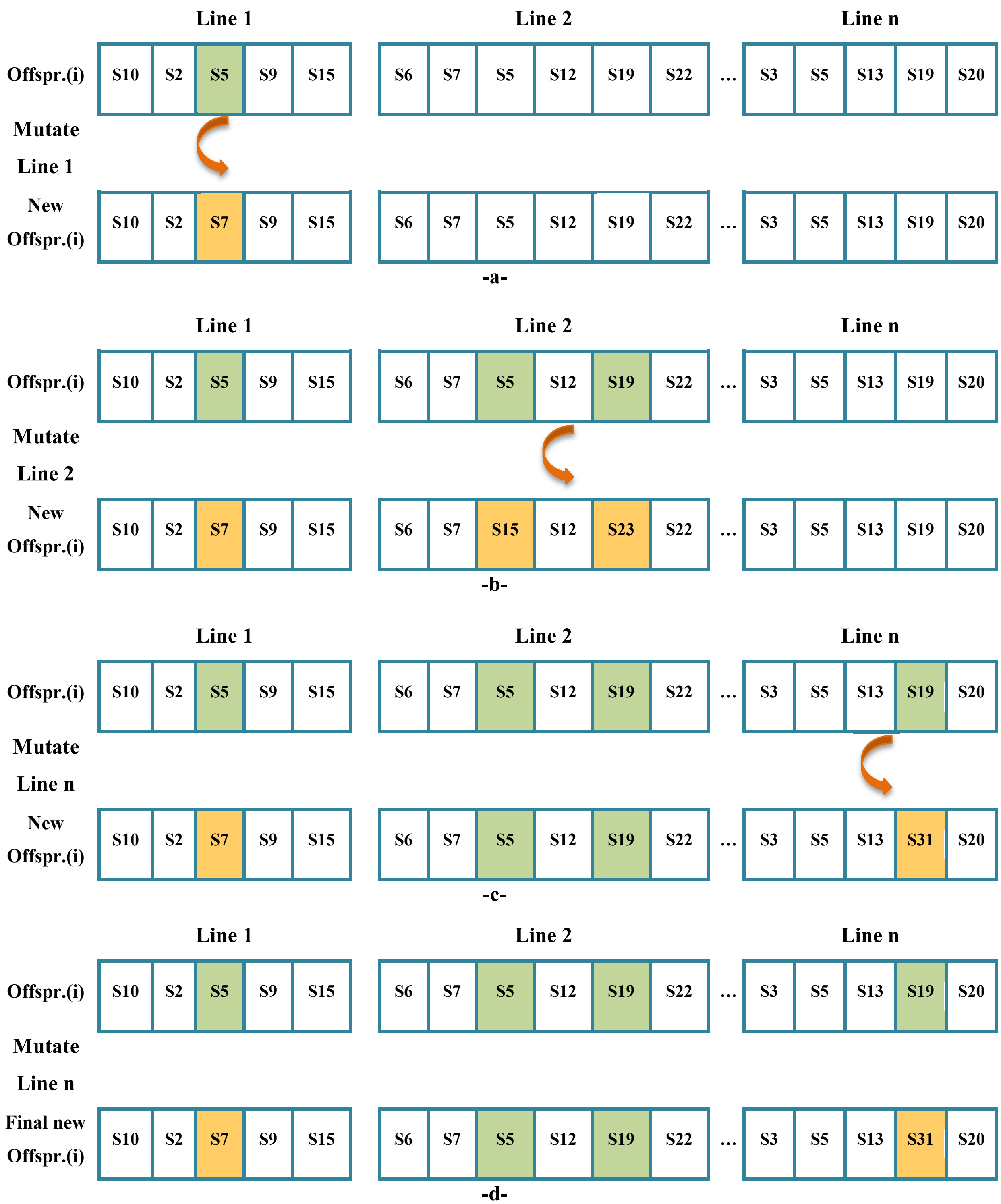


Figure 5.17: Illustration of Mutation at Gene Level Operator

5.5.3.3 Mutation at Gene Bit Level (MGBL)

This operator is designed to take individuals from the offspring pool and randomly mutate them at the gene bit level (station) to produce new offspring. It individually selects a number of gene bits of the individual's chromosome with probability Pm, and replaces them with randomly selected gene bits to produce a new offspring. The main difference between this operator and the two mutation operators above, MCL and MGL, is that it mutates the individual at the gene bit (station) level while the others mutate the individual at the chromosome (line network) and gene (individual line) levels respectively. The following key steps detail the procedure of this operator:

- **Step 1:** Set two counters, one for counting the total number of mutations (Mtm) that occur over the total number of generations, and the other for counting the new feasible offspring (Mtf) produced from mutating the individuals within the offspring pool, i.e., set Mtm= 0 and Mtf=0.
- **Step 2:** Select an individual from the offspring pool produced by the crossover operator.
- **Step 3:** Successively determine the gene bits of the selected offspring to be mutated. This is achieved by successively visiting the corresponding gene bit of the offspring and generating a random number between 0 and 1. If the generated number is greater than Pm, proceed to the next gene bit. Otherwise, assign the gene bit to be mutated and then proceed with the following steps.
- **Step 4:** Replace the gene bit identified for the mutation in step 3 with a randomly selected gene bit to produce a new offspring as illustrated in figure 5.18a, and increase the Mtm counter by one (i.e., Mtm= Mtm+1). Then check the feasibility of the generated new offspring with respect to the required constraints identified in section 3.4.2.2, and apply the following if-else scenario.
  - If the generated new offspring satisfy the required constraints, increase the Mtf counter by one (i.e., Mtf = Mtf +1) and go back to step 3 to mutate the next gene bit. It should be noted that, apart from the mutated gene bit, the other gene bits of the produced new offspring are copies of the corresponding offspring gene bits, as illustrated in figure 5.18b, 5.18c and 5.18d.
  - Else (i.e., if the generated new offspring violate the constraints) reject the gene bit mutation and return the mutated gene bit back to its initial state before the mutation, and then go back to step 3 to mutate the next gene bit (see figure 5.18b and 5.18c).

- **Step 5** repeat steps 3 and 4 until all the corresponding gene bits of the selected individual are mutated and the final new offspring is produced (see figure 5.18d).
- **Step 6:** repeat steps 2 to 5 until all the individuals within the offspring pool are mutated.
- **Step 7:** Calculate the success rate (Msr) by dividing the value of Mtf by the value of Mtm. It should be noted that this value is calculated after completing the required number of generations.

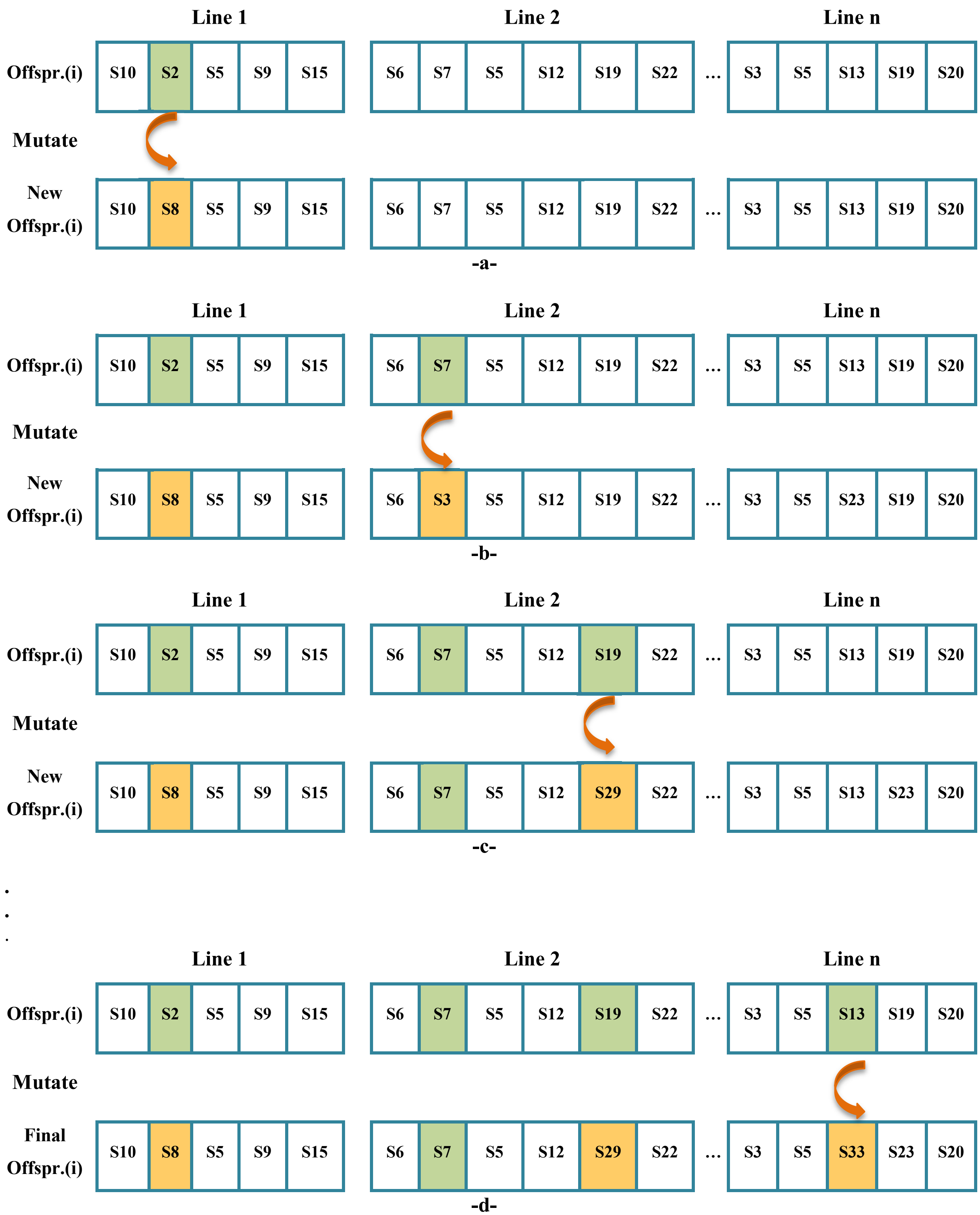


Figure 5.18: Illustration of Mutation at Gene Bit Level Operator

5.6 Determination of the Optimal Solution

As depicted in figure 5.10, after the selection, the crossover and mutation operators are applied to the initial population so that a new population is formed for the next generation. In a typical GA the population size is constant over the whole generations, and therefore, the proposed GA checks the size of the offspring population produced from applying the crossover and mutation operators at each generation. If it is the same as the original population size the entire original population is replaced with the generated offspring for the next generation. If the size of the offspring population is smaller than the original population, the worst individuals in the original population are replaced with the generated offspring , and new individuals generated randomly using the procedure described in section 5.3 to obtain the required population size for the next generation. The percentage of new individuals inserting into the new population is assumed to be user defined. The higher the percentage of new individuals injected into the new population, the higher the exploration of the search would be, and thus the lower the probabilities of yielding a sub-optimal solution. In addition, replacing a portion of the worst individuals with the generated offspring, which is known as 'elitism' in GA, guarantees that the average fitness values of the population's individuals will not decrease from one generation to another and thus makes the algorithm to converge on the optimal solution faster.

Unfortunately, since GAs cannot be expected to terminate on the global optimum solution (i.e., the best solution) spontaneously, and since they are also not guaranteed to find the global optimum solution, this process of selection, crossover and mutation has to be stopped at some point according to predefined criteria. In general, two stopping criteria are employed in GAs. The first is the improvement of the objective function, i.e., the evolution process will be stopped if the improvement in the objective function is less than a specified value. The second is the maximum number of generation, which is fixed a priori, i.e., the evolution process will be stopped if the predefined number of generation reaches its value. In this thesis, if one of the following two criteria is satisfied the evolution process is terminated and the best individual in the current population is identified as the optimal solution.

1- Terminate if the best individual of the population is not further improved over a specified number of successive generations, which is defined by the potential user/planner.
2- Terminate if the number of generations, which is also defined by the potential user/planner and supposed to be large enough to efficiently handle the whole search process, reaches its value.

5.7 Summary

This chapter presents an integrated optimization algorithm to simultaneously determine the optimal locations of stations and the associated line network connecting the stations. This algorithm aims to resolve the essential trade-offs between an effective rail system that provides high service quality and benefits for both the passenger and the whole community, and an economically efficient system with acceptable capital and operational costs. To achieve this, it minimizes total system cost, which is a linear function of passenger, operator and community cost. Through the passenger cost function, the algorithm tries to deliver adequate service quality for railway passengers by either decreasing passengers' rail travel time compared to bus and car travel time, or by increasing railway usage, or a combination of the two. Through the operator cost function, the algorithm tries to increase the railway system revenue gain and achieve an economically efficient system operation by increasing railway usage. Through the community cost function, which covers all the construction cost elements, the algorithm tries to minimize the construction costs. The proposed integrated optimization algorithm therefore, resolves the trade-offs by trying to increase the number of railway users in order to reduce both the proportional passenger and operational costs, while not significantly increasing the construction costs. This implies that the algorithm may increase the system construction costs to a certain extent to accommodate more railway patronage, while on the other hand reducing passenger and operator costs by attracting more car and bus users to switch to rail.

The optimization algorithm is based on Genetic Algorithms (GA) with the support of a GIS. In the first step, the algorithm generates a set of initial alternative solutions by selecting a set of stations from the candidate station pool identified as described in chapter Four, as well as generating a network of lines connecting the selected stations while satisfying various general requirements of rail transit system planning. In the second step, it evaluates the fitness of the generated alternative solutions with respect to the mathematical functions of the three stakeholders' requirement criteria; passenger, operator and community cost. The algorithm incorporates various local factors, such as land use patterns, land value, topography and soil conditions into the evaluation framework via interaction with the GIS, and embeds a demand forecasting module for predicting the usage of the proposed rail transit system. In the third step, the algorithm evolves the generated initial solution over a series of generations/iterations based on their fitness values to converge towards the optimal solution, applying three genetic operators, selection, crossover and mutation. The evolution process is continued until the convergence criteria are met and the optimal solution is determined in step four.

To demonstrate the algorithm's effectiveness in finding a robust solution and to reveal its efficiency in regions with complex topographies, it is applied to a real world case study, the City of Leicester (UK) in the next chapter. An extensive numerical study is also included to reveal the importance of screening the study area for feasible station locations prior to the optimization process on the performance of the optimization model. This is in addition to examining the proposed different parameters and structures of the GA operators to draw conclusions on the best parameters and structures for the solution of the problem rail transit system planning.

Chapter 6 – Case Study and Sensitivity Analysis

Chapter 3 presented a practical method for planning a rail transit system that aims to help planners to determine the optimal locations for rail stations and the associated line network. This new method incorporates the complex correlations and interactions between the rail line alignments and station locations into an integrated optimization model while taking into account the various requirements and constraints inherent in rail transit system planning. The optimization model starts by screening the study area for potential station locations with respect to the various requirements of passengers, operators and the community. It then uses these potential stations as a pool from which it selects the optimal set of stations while simultaneously generating the best line network connecting the selected stations using a heuristic solution algorithm based on GA and a GIS database.

The details of how the proposed optimization model screens the study area for potential station locations within the context of a GIS-based algorithm is presented in chapter 4 with the help of a case study on the City of Leicester (UK). This is followed by the illustration of the design and formulation of the proposed heuristic optimization algorithm in chapter 5. This chapter applies that proposed integrated optimization model to the Leicester case study to examine its effectiveness in finding a robust solution in large scale regions with complex topographical features.

Section 6.1 presents the case study and preparation of the required datasets. Section 6.2 presents the application of the proposed integrated optimization model to the case study. This section also demonstrates the practical applicability of the model, not only for planning a new rail system but also for expanding an existing rail transit system. Section 6.3 presents a sensitivity analysis to examine how the proposed optimization model interacts with variations in travel demand by adjusting its station selection and associated line alignment to minimize the total system cost. Section 6.4 compares individual and simultaneous optimization methods for planning a rail transit system to determine the benefits of the latter. Section 6.5 demonstrates the importance of performing a feasibility analysis for potential station locations prior to the optimization process in facilitating an efficient search process towards the optimal solution. This section also investigates the effect of different threshold values for the requirement criteria on the optimization search behaviour. Section 6.6 examines the different parameters of the proposed genetic algorithm to determine the setting that best fits the framework of the rail transit system planning problem. Section 6.7 evaluates the goodness of the solutions obtained by the new optimization model. A summary of the work and findings is presented in section 6.8.

6.1 Case Study and data pre-processing

This section exploits the real world case study used in chapter 4, City of Leicester, to further examine the effectiveness and practical applicability of the proposed integrated optimization model. The required datasets for the study area were obtained from several agencies including Leicester City Council, the Ordnance Survey of Great Britain, the British Tunnel Society and the Land Registry. The datasets included the City of Leicester's geographical boundary, land use patterns, census data, land values, topography and geological maps. The required input data were extracted from these datasets and retrieved in ArcGIS database in order to apply the optimization model. The preparation of some of the required model input data, which were used for determining potential locations of stations, is presented in detail in section 4.3.1. This section focuses on the preparation of the additional input data specifically geological, topographical and origin-destination trip matrix data.

1- **Geological data :** this dataset was obtained from the Ordnance Survey of Great Britain in the feature class format in ArcGIS. The geometric shape of this dataset is polygonal. Each polygon, which represents a particular portion of the study area, has the attributes of corresponding geological properties and composition. However, this dataset has no information on the vertical variation of the ground composition (i.e., vertical variation in the thickness of soil and rock bed layers). To obtain such information, requires resource intensive soil investigations, not available for this thesis. . Therefore, the data on the depths of soil and rock bed layers are assumed, on the basis that it should be sufficient to examine the effectiveness of the new optimization model. As detailed in section 5.4.3, the proposed optimization model uses this data to calculate both the tunnel and escalator barrel costs.

2- **Topographical data:** this dataset was also obtained from the Ordnance Survey of Great Britain in the feature class format in ArcGIS. The data consists of polyline features representing the contour lines of the study area. The contour lines are at 5 metre vertical intervals. The proposed optimization model integrates this data with the geological data to calculate both tunnel and escalator barrel costs as explained in section 5.4.3.

3- **Origin-Destination trip matrix data:** this data was obtained from Leicester City Council, and was generated from the Leicestershire Integrated Transport Model (LLITM). The dataset consists of the origin-destination trip matrix at zonal level for each alternative transport mode in the city (i.e., car, bus and active mode) at four different time periods (AM, IP, OP, and PM) of a day for the year 2015. For simplicity, this thesis assumes that the trip distribution pattern between the traffic analysis zones of the study area does not change after building the proposed rail system. Therefore, for each particular zone pair, the total demand matrix was computed by summing up the corresponding demand matrices of the three existing transport modes. The proposed optimization model uses this demand matrix as an input and splits it into four matrices, one for each mode (rail, car, bus and active mode), via the mode choice model embedded within it, as discussed in section 5.4.4.

6.2 Model Results

Using the datasets prepared in section 4.3.1 and 6.1, and the parameters presented in table 6.1, the proposed optimization model was implemented. It should be noted that the model is designed to incorporate flexibility so that it can be used both for planning new rail systems and expanding existing ones. To examine the effectiveness of the model for each of these potential uses, two planning scenarios were undertaken. The first scenario applies the model for planning of a new rail system while the second scenario employs the model to expand a predefined existing rail system. The following two subsections detail the results of these two scenarios.

Table 6.1: Values of the Input Parameters Used in the Case Study

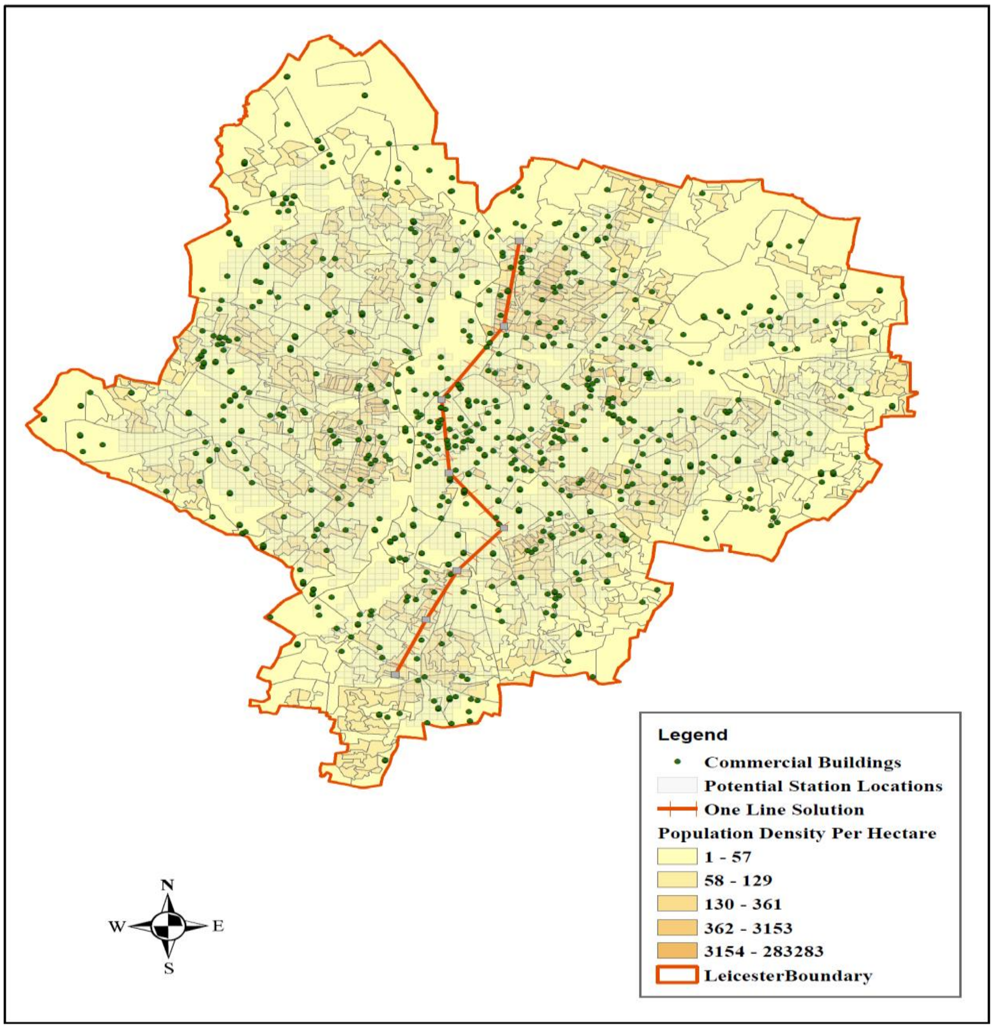
Parameters	Description	Values
V_t	train travel speed	80 km/hr
V_a	passenger walking speed	4 km/hr
H_w	train headway at peak/off peak time period	10/15min
A_0	unit cost of access time	£14/hr
W_0	unit cost of waiting time	£14/hr
T_0	unit cost of in train travel time	£7/hr
S_0	unit cost of time spent searching for parking space	£14/hr
M_{t0}	unit operation and maintenance cost for train	£0.21/passenger-Km
M_{c0}	unit operation and maintenance cost for car	£0.26/passenger-Km
M_{b0}	unit operation and maintenance cost for bus	£0.16/passenger-Km
S_b	costs of building and equipping the stations with necessary facilities	£10^6 / station
C_{Tr}	unit cost of track	£650/m
Ns_{min}	minimum number of stations along each rail line	4 stations
Ns_{max}	maximum number of stations along each rail line	7 stations
Δs_{min}	minimum spacing between stations	800 m
Δs_{max}	maximum spacing between stations	1500 m
Pn_{min}	minimum population coverage rate by the rail line	50%
Cn_{min}	minimum number of common/transfer stations along each rail line	1
On_{max}	maximum overlap rate between the rail lines	30%

Parameters	Description	Values
φp	coefficient of passenger cost	1
φo	coefficient of operator cost	1
φc	coefficient community cost	1
Pn	GA parameter-Population size	150
Pc	GA parameter-Crossover probability	0.7
Pm	GA parameter-Mutation probability	0.3
Gen	GA parameter-Number of generations	50

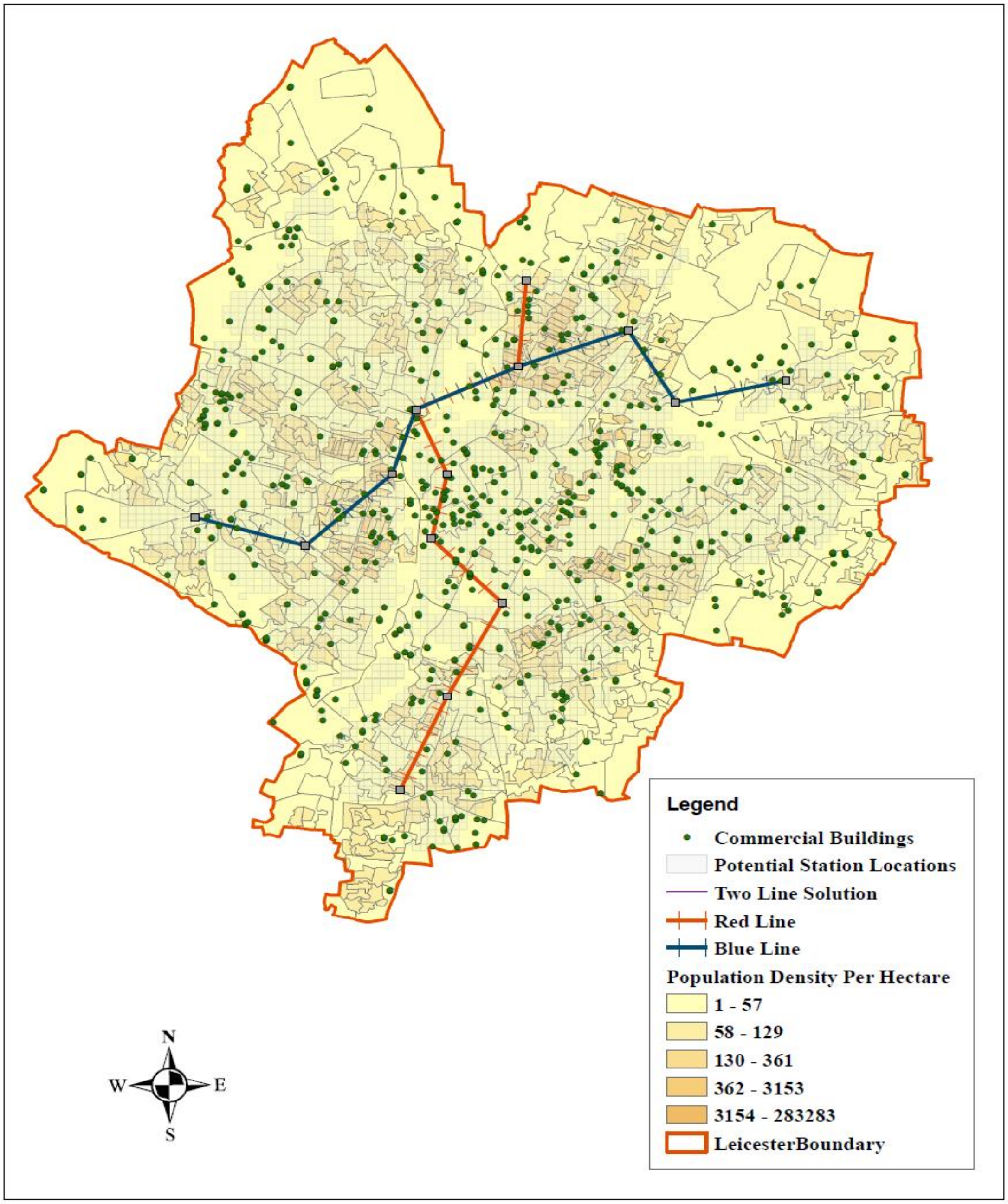
6.2.1 Planning a New Rail Transit System

As discussed in section 5.3, it is assumed that the desired number of rail lines and terminal stations for each line are determined by rail transit planners depending on the major traffic flow pattern of the area/city to be served. Accordingly, the proposed model generates the corridors within which the line network and associated stations are to be located based on the terminal stations, as illustrated in figure 5.2. The basic topological configuration of the system is therefore, bounded by these corridors. The range of topological configurations that can be considered in this thesis are not limited, in the sense that the proposed optimization model can accommodate any basic topology and find the solution that best fits the desired objectives.

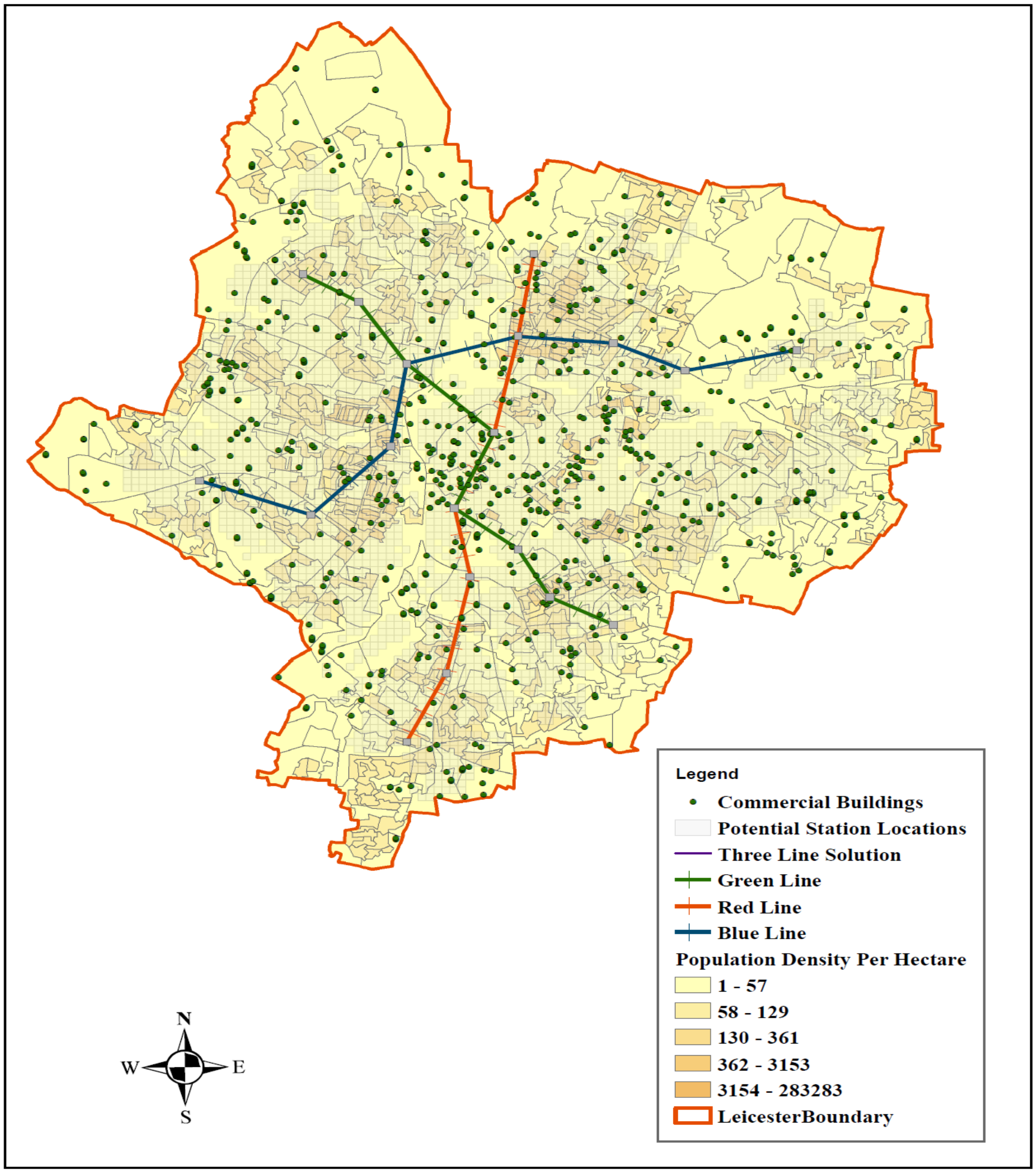
To examine the ability of the model to handle the different topological configurations of a rail transit system, three different scenarios were applied. The first scenario considers a proposed rail system consisting of a single rail line, while the second and the third scenarios consider rail systems consisting of two and three lines, respectively. It should be noted that the terminal stations of every line in the three scenarios were determined based on the origin-destination trip matrices obtained from Leicester City Council; these matrices represent traffic flow patterns, knowledge of which is crucial for strategic planning and the management of transportation networks. Figures 6.1a to 6.1c show the results of these three test scenarios. In the first scenario, the proposed rail line links south and north of the city. In the second scenario, two rail lines were proposed, one linking the south and north of the city and the other one linking the south-west and north-east. The three proposed rail lines in the third scenario link south and north, south-west and north-east, and north-west and south-east of the city. Throughout the rest of this chapter these lines are referred to as the Red Line (south/ north); Blue Line (south-west/ north-east) and Green Line (south-east/ north-west) (see figure 6.1).



(a)


Legend
Commercial Buildings
Potential Station Locations
Two Line Solution
Red Line
Blue Line
Population Density Per Hectare
1 - 57
58 - 129
130 - 361
362 - 3153
3154 - 283283
LeicesterBoundary
N
W
E
S

(b)



(c)

Figure 6.1: Optimized Station Locations and Associated Line Network: (a) One Rail Line Solution; (b) Two Rail Line solution; (c) Three Rail Line Solution

The results shown in figure 6.1 reveal that the model can efficiently generate different topological configurations. Figure 6.1b, exhibits a star configuration while the proposed solution in figure 6.1c exhibits a combination of star and triangle configurations. It is also notable that the line alignments and associated station locations change as the overall size of the system changes to minimize the total system cost while also satisfying the required constraints. For example, compared to the first scenario, the alignment and station locations of the Red Line are different in both the second and third scenarios. Specifically, comparing the second scenario with the first, four stations starting from the fourth station in the Red Line (northbound from the southern terminus) have shifted towards the western side of the city centre, intersecting the Blue Line to the north-west of the city centre in the second scenario. Accordingly, the overall alignment of the Red Line was adjusted towards these shifted stations, as illustrated in figure 6.1b. Similarly, the Red and Blue Line alignments and associated station locations have changed in the third scenario compared to the second scenario. These results confirm that the proposed model is able to deal effectively with different system sizes when optimizing station locations and associated line alignments. It should be noted that although the model takes into account some geometrical requirements for line alignment, such as minimum and maximum gradient, as explained in section 5.4.3.2.2, it disregards other geometric requirements for horizontal and vertical rail line alignments for the sake of computational simplicity. Comprehensive consideration of various geometric requirements for both horizontal and vertical rail line alignment design (such as minimum radius of horizontal curves, minimum tangent lengths, minimum super elevation rates and minimum and maximum radius and length of vertical curves) is essential for passenger comfort and safety as well as smooth and safe operation of railway vehicles. Therefore, addressing such aspects in future works is important.

The optimized total system cost for each of these three solutions is presented in table 6.2. The results show that this decreases significantly with the size of the proposed rail system solutions. For example, the total system cost for the single rail line system, i.e. the Red Line only, was -46.514 M£, but this decreased to -203.866 when the size of the proposed rail system is enlarged to accommodate three rail lines. It is notable that the community cost, which involves the construction cost of the system, increased while both passenger and operator costs decreased significantly. The passenger and operator cost decreased mainly due to more passenger being attracted to use the rail system, the daily rail passenger number increased by about four times when the size of the proposed system enlarged to accommodate three lines, as depicted in table 6.2 These results confirm that the model resolves the essential trade-off between the community cost and both passenger and operator costs. On the one hand, it increases the community cost of the system to accommodate and attract more railway passengers. On the other hand, it reduces both passenger and operator costs by shifting more trips from car and bus to rail.

It is important to note that the sign of both passenger and operator costs is negative. For the passenger costs this signifies that the cost of passengers' travel time has decreased, or been saved, by using the proposed rail system instead of using car or bus transport modes (see equation 3.10). As for the operator cost, the negative sign refers to the total cost of operations and maintenance decreasing due to the use of the proposed rail system instead of using either car or bus transport modes (see equation 3.11).

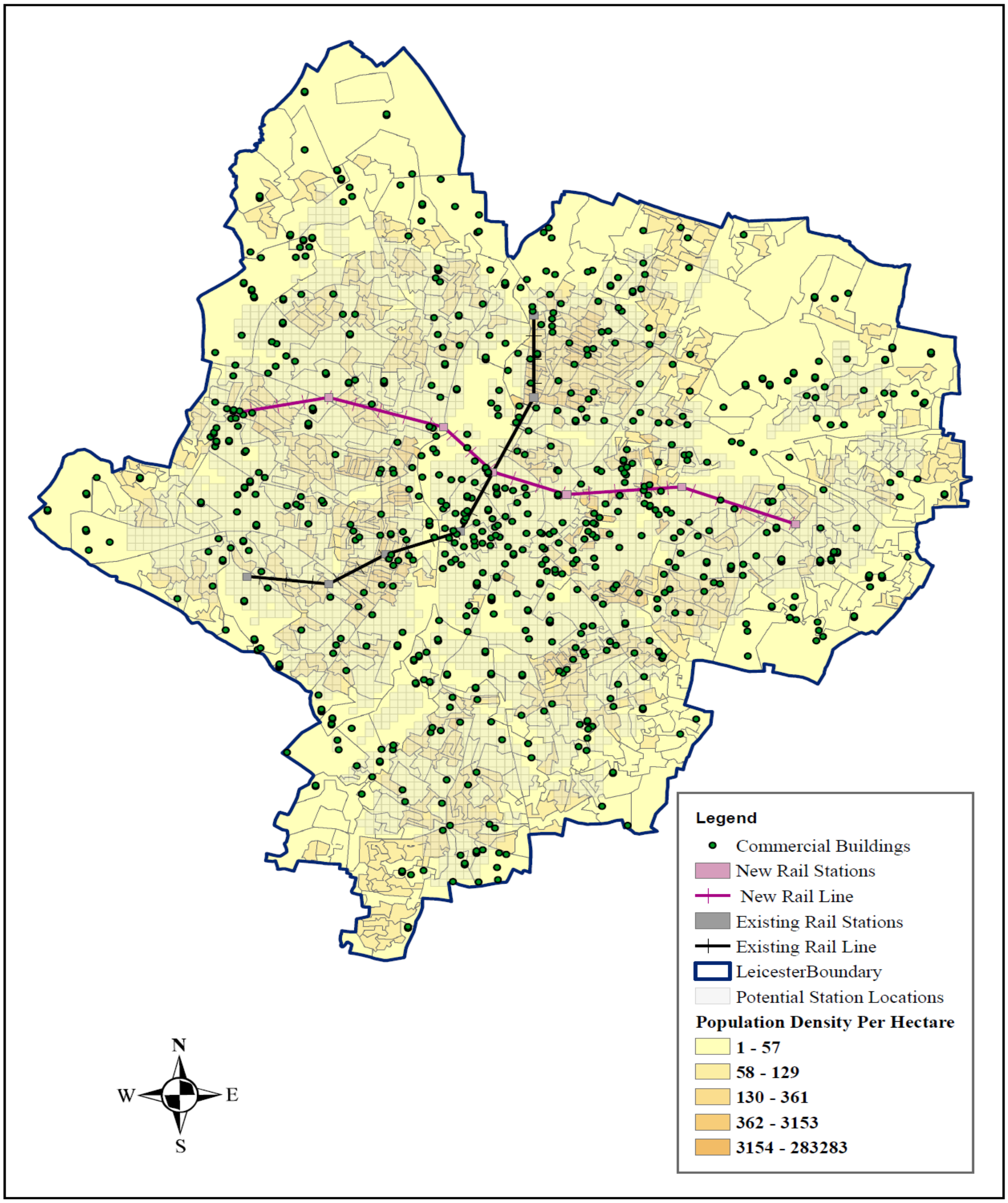
Table 6.2: Optimized Costs of the Three Planning Scenarios

Scenarios	Passenger Cost (M£)	Operator Cost (M£)	Construction Cost (M£)	Total System Cost(M£)	Line Network Length (km)	Total number of Stations	Daily Passengers
One Line	-144.750	-33.155	131.489	-46.416	8.002	6	5289
Two Lines	-304.305	-84.132	253.368	-135.069	18.131	16	16715
Three Lines	-441.808	-144.422	382.602	-203.628	23.903	23	20307

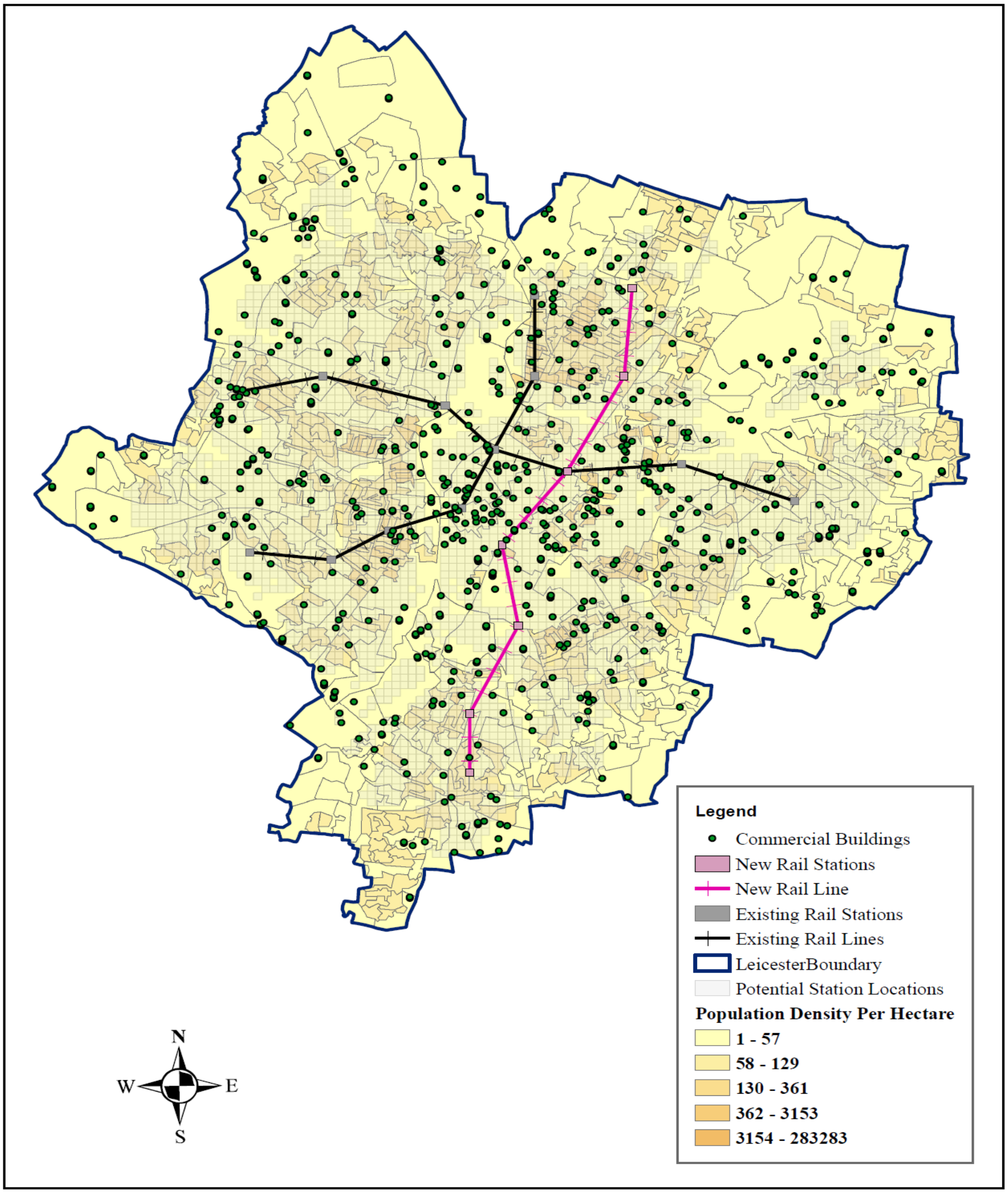
It is also noteworthy that the number of lines that can be generated by the model for connecting the stations is not limited. That is, it can generate an unlimited number of lines to connect stations, but obviously the higher the number of lines the longer the running time of the model will be. In addition, the results reveal that the generated solutions mainly cover the most densely populated and important activity zones of the city, indicating the effectiveness of the model in finding solutions that fit the framework of the rail transit system planning requirements.

6.2.2 Expanding an Existing Rail Transit System

The effectiveness of the proposed integrated optimization model in expanding an existing rail system is tested in this section by considering two different scenarios. The first scenario is intended to expand an existing rail system from a single line to two lines, while the second scenario is intended to expand an existing rail system from two lines to three lines. It should be noted that the existing rail systems defined in this section are hypothetical, and were created only for undertaking these two scenarios. Obviously, in the cases where the proposed model is used for expanding an existing rail system, comprehensive information on the existing system must be made known before implementing the model. This information, which is used as an input for the proposed model, involves the geographical locations of the stations and line(s), the ground elevation of the stations and line(s) and the depth of the associated tunnels. Figure 6.2a and 6.2b illustrate the results of these two scenarios respectively.


Legend
Commercial Buildings
New Rail Stations
New Rail Line
Existing Rail Stations
Existing Rail Line
LeicesterBoundary
Potential Station Locations
Population Density Per Hectare
1 - 57
58 - 129
130 - 361
362 - 3153
3154 - 283283
N
W
E
S

(a)



(b)

Figure 6.2: Expanding an Existing Rail Transit System: (a) Addition of one line to an existing line; (b) Addition of one line to two existing lines

The results show that the proposed model is effective not only in expanding existing rail systems with a single rail line but also rail systems with multiple rail lines and complex topological configurations. Furthermore, the results show that the existing rail lines are connected with the proposed new lines without changing their original alignments and associated station locations. This means that the proposed model optimizes the new rail line alignment and station locations to minimize the total cost of the system while satisfying the required constraints identified in section 3.4.2.2 and making sure that neither the alignment(s) nor station locations of the existing rail line (s) are changed. Table 6.3 presents the optimized total system cost for these two scenarios.

Table 6.3: Optimized Cost when Expanding an Existing Rail Transit System

Scenarios	Cost Items	Rail System before Expansion	Rail System after Expansion
1	Passenger Cost (M£)	-102.190	-196.604
	Operator Cost (M£)	-14.446	-86.129
	Community Cost (M£)	98.887	201.045
	Total Cost (M£)	-17.749	-81.688
	Daily Passenger	1772	5819
	No. of Rail Station	7	14
	Total Line Length (km)	5.872	13.131
2	Passenger Cost (M£)	-196.604	-339.732
	Operator Cost (M£)	-86.129	-167.699
	Community Cost (M£)	201.045	315.654
	Total Cost (M£)	-81.688	-191.777
	Daily Passenger	5819	16125
	No. of Rail Station	14	21
	Total Line Length (km)	13.131	20.888

As depicted in table 6.3, in both scenarios, the total cost of the proposed rail systems decreased significantly after the expansion. For example, when the existing rail system expanded from one line to two lines (scenario 1) the total system cost decreased by 63.939 M£ from (-17.749 M£ to -81.688 M£). This reduction can be explained mainly by the significant increase in the number of daily rail passengers. Passenger numbers more than doubled after the new line was added to the system, which, in turn, resulted in reduced passenger and operator costs. These reductions in the total system costs reflect the degree to which the expansion of the existing rail system benefits the rail transit system stakeholders. For example, with addition of a third rail line (scenario 2) the passengers' travel time cost and operation and maintenance cost that will be saved or reduced due to increased rail use compared to other transport modes (car and bus) reached 339.732 M£ and 167.699 M£ from 196.604 M£ and 86.129 M£ respectively over the life span of the rail system.

It is important to note that, since the alignment and station locations of the existing rail line(s) were restrained to remain unchanged while the alignment(s) and station locations of the new line(s) are optimized, their respective community costs remained constant during the optimization process.

6.3 Impact of Demand Variation on the Optimization Results

This section examines how the proposed integrated optimization model interacts with variations in demand distribution by adjusting station locations and the line alignments linking the stations to minimize the total system cost. This is achieved by increasing the total demand from/to the six TAZs of 5001, 5002, 5003, 5005, 5007 and 5027 (figure 6.3) by a factor of five, and comparing the optimized station locations and associated line network alignments of the actual and the adjusted demand distributions. The proposed Red Line in section 6.2.1 is used for this test. In general, the line alignment and the locations of most of the stations are changed, particularly those stations which are located close to the TAZs with the changed demand. Three of the six stations selected for the actual demand distribution are shifted to locate very close to, or within, the six TAZs with the adjusted demands, as shown in figure 6.3. Also the line alignment through the first intermediate station from the south terminus is adjusted towards the shifted station locations.

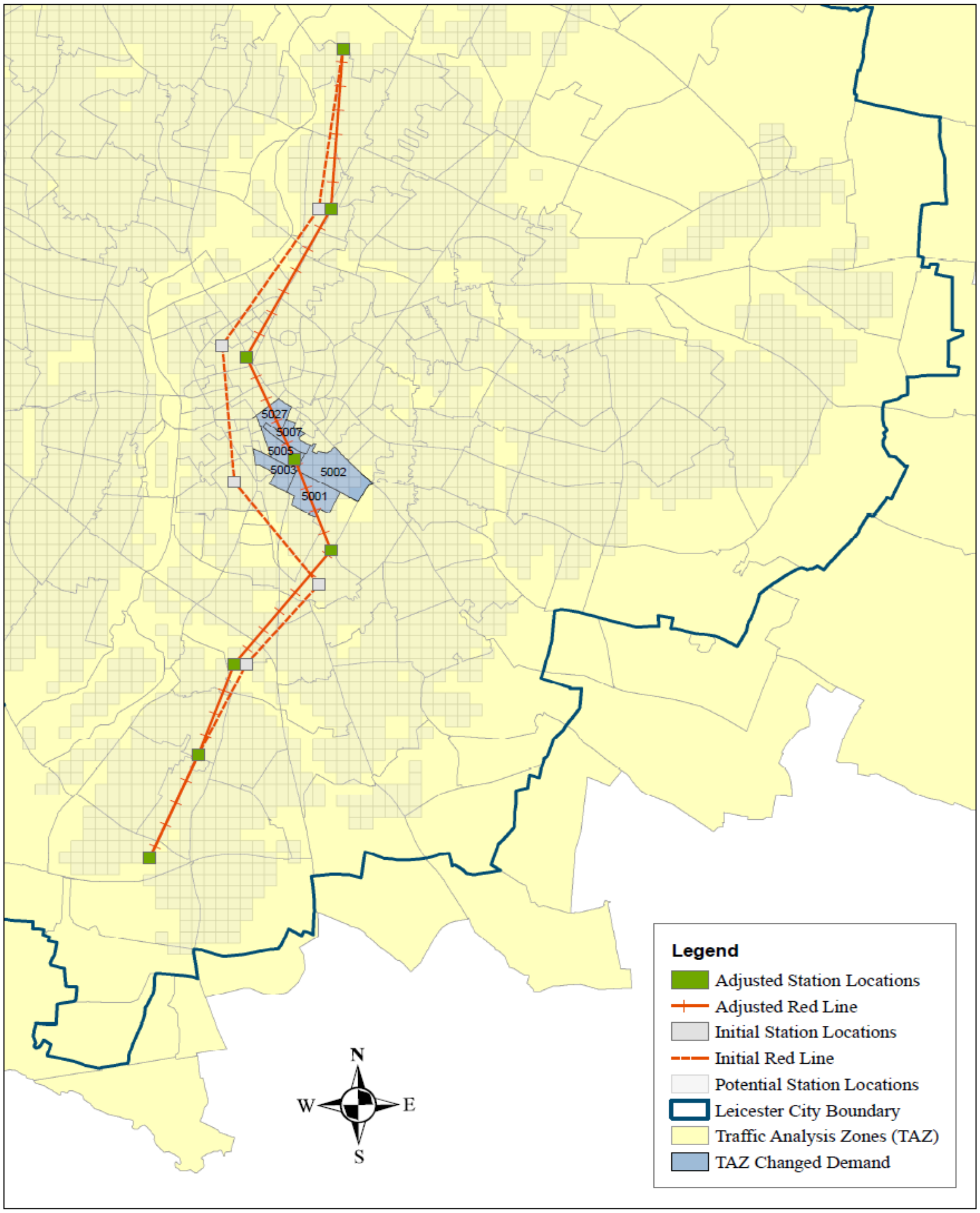


Figure 6.3: Illustration of the Impact of Demand Variations on the Optimization Results

Compared to the optimized solution generated for the actual demand distribution, the optimized solution generated for the adjusted demand incurred an increase in the community cost of 3.382 M£ from 131.489 M£ to 134.871 M£. However, as shown in table 6.4, the passenger and operator costs decreased by 10.536 M£ and 14.214 M£ respectively for the adjusted demand solution. This is mainly due to attracting more than 17% of the original demand by directly serving the TAZs with higher demand. These results confirm that the proposed model is effective in recognizing travel demand patterns and accordingly optimizing station locations and the associated line network. In addition, these results prove that the model is effective in making the trade-off between the minimization of construction costs and the maximization of rail usage (passenger and operator costs). On the one hand, it increases the construction cost of the system to accommodate further rail usage while, on the other hand, reduces both passenger and operator costs by attracting more trips from car and bus to rail.

Table 6.4: Optimized Cost for Actual and Adjusted Demand

Scenarios	Passenger Cost (M£)	Operator Cost (M£)	Construction Cost (M£)	Total System Cost(M£)	Line Network Length (km)	Total number of Stations	Daily Passengers
Actual Demand	-144.750	-33.155	131.489	-46.42	8.002	6	5289
Adjusted Demand	-155.285	-47.369	134.871	-67.78	7.805	6	6221

6.4 Comparison of Individual and Simultaneous Optimization

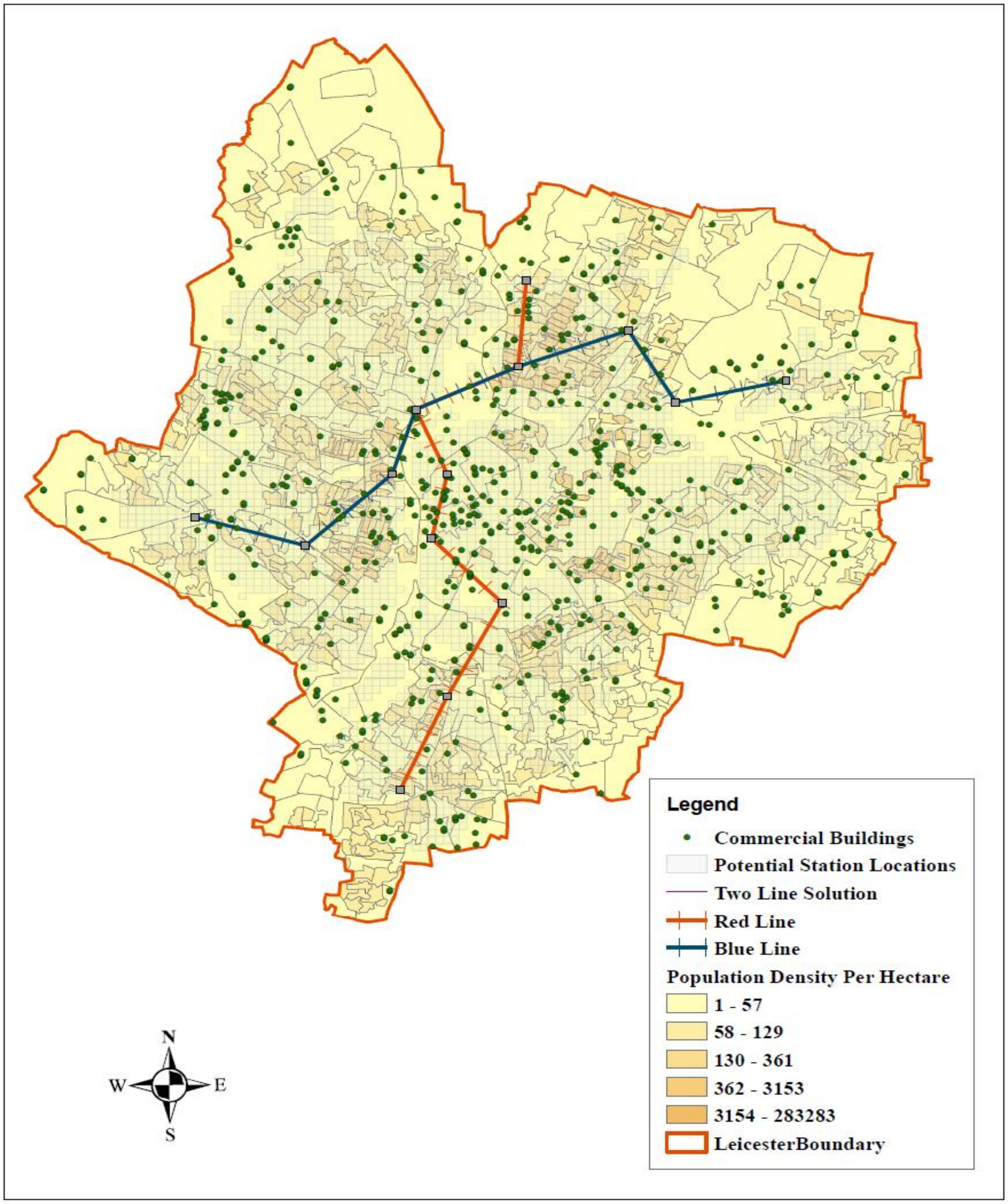
This section compares independent and simultaneous optimization methods for planning a rail transit system. The former optimizes alignment and associated station locations of each line in the rail system individually, while the latter optimizes the alignments and associated station locations of the all lines in the system simultaneously. Table 6.5 presents the optimization results of these two methods on a rail transit system with two rail lines.

Table 6.5: Comparison of Independent and Simultaneous Optimization for a Two Rail Line System

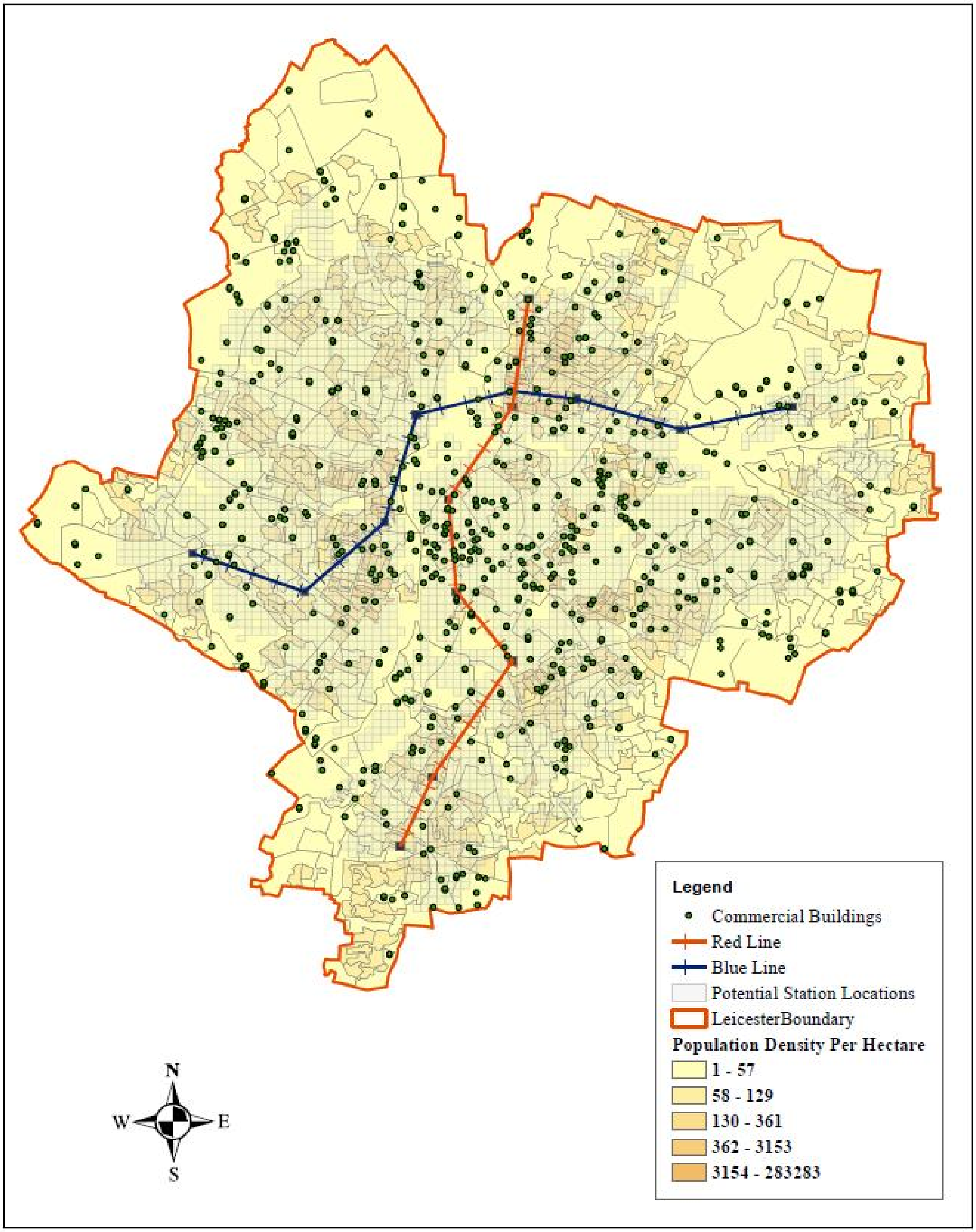
Two Lines (Red-Blue)Lines	Simultaneous Optimization	Individual Optimization			Difference %
		Red Line	Blue Line	Total	
Passenger Cost (M£)	-304.305	-144.750	-116.975	-261.725	-16.27
Operator Cost (M£)	-84.132	-33.155	-41.290	-74.445	-13.01
Community Cost (M£)	253.368	131.489	125.211	256.700	1.30
Total Cost (M£)	-135.069	-46.416	-33.054	-79.470	-69.96
Line Network Length (km)	18.131	9.123	7.988	17.111	-5.96
Total number of Stations	14.000	7.000	8.000	15.000	6.67
Daily Passenger	16715	5289	3874	9163	-82.42

Compared to the individual optimization, simultaneous optimization significantly reduced the total cost by about 70% from -79.470 M£ to -135.069 M£. The daily passenger numbers for the system created by simultaneous optimization were 82.42% more than those using the system created through individual optimization. Also, the simultaneous optimization decreased the community cost from 256.7 M£ to 253.368 M£, passenger cost from -261.72 5M£ to -304.305 M£ and operator cost from -74.445 M£ to -84.132 M£. These numerical results prove the advantage of simultaneous optimization over individual optimization.

Figure 6.4a presents the solution generated from the simultaneous optimization and figure 6.4b presents the solution generated from the individual optimization. In general, both solutions exhibit a similar star-shaped topological configuration. However, the length of the total line network, and the station numbers, as well as station locations and line alignments generated from the two optimization methods, are different. The total line network length and station numbers generated from the simultaneous optimization were 18.131 km and 15 stations, while with the individual optimization they were 17.111 km and 16 stations. It is important to note that the locations of the first intermediate stations from the terminal stations did not change for either of the optimization methods while the locations of the other intermediate stations and their associated line alignment changed. The stations in the individual optimization and the associated line alignment were generated to attract the maximum possible number of rail passengers for each line individually without considering the availability of other lines in the system and possible passenger transfer between these lines. In contrast, the simultaneous optimization method selected the station locations to attract the maximum possible number of railway passengers for all the lines in the system, taken together and considering the connectivity between the lines. As depicted in figure 6.4a, the generated rail lines intersect at two different stations in the north-west and northern sides of the city centre. This, in turn, resulted in a significant increase in passengers attracted to use the system, and subsequently decreased the total cost significantly compared to the solution generated from the individual optimization method, where the lines, although cross, do not intersect at a station..


Legend
Commercial Buildings
Potential Station Locations
Two Line Solution
Red Line
Blue Line
Population Density Per Hectare
1 - 57
58 - 129
130 - 361
362 - 3153
3154 - 283283
LeicesterBoundary
N
S
E
W

(a)



(b)

Figure 6.4: Optimization Methods of a Two Rail Line System: (a) Simultaneous Optimization Method; (b) Individual Optimization of a Two Rail Line System

6.5 Impact of the Determination of Feasible Station Locations Prior to the Optimization Process

This section examines how screening the study area for potential station locations prior to the optimization process affects the performance of the proposed model in finding a good solution. Furthermore, it investigates the effect of different threshold values for the requirement criteria in terms of narrowing down the optimization search space to determine values that would not result in the exclusion of potential optimal station locations. This was achieved by using five different threshold values for the requirement criteria, specifically 30th, 40th, 50th 60th and 70th percentiles, in order to identify potential station locations and then to compare the optimized total system cost obtained for each of these threshold values with each other. The reason for not including the 10th and 20th percentiles in this section was because of the marginal difference between these two threshold values and 30th percentiles in terms of the number of feasible station locations, as depicted in figure 4.12. Table 6.6 presents the optimized cost for different threshold values of the requirement criteria for the two rail line system proposed in section 6.2.1.

Table 6.6: Comparison of Different Threshold Values of the Requirement Criteria on the Optimized Cost

Threshold Values of the Requirement Criteria	Passenger Cost (M£)	Operator Cost (M£)	Community Cost (M£)	Total Cost (M£)	Network Length (km)	No. of Stations	Daily Passengers
30th Percentile	-262.933	-91.516	233.086	-121.363	-262.933	15	11602
40th Percentile	-309.767	-77.4735	256.355	-130.885	-309.767	16	16545
50th Percentile	-304.305	-84.132	253.368	-135.069	-304.305	16	16715
60th Percentile	-305.984	-104.567	255.731	-154.820	-305.984	16	17685
70th Percentile	na	na	na	na	na	na	na

The results show that the proposed optimization model can find a better solution with higher threshold values, up to a point. The total cost of the system decreases by 27.56% from -121.363 M£ to -154.820 M£ when the threshold values of the requirement criteria is increased from the 30th to the 60th percentiles, but compared to these two extremes the reduction in the total cost of the system between the 40th and 50th percentiles is low, decreasing by just 3.19% from -130.885 M£ to -

135.069 M£. This is because of the relatively low difference in the number of potential station locations with these two threshold values. The numbers of potential station locations were 3794 and 4439 for the 40th and 50th percentiles respectively. The number of potential station locations was 5125 and 2414 for 30th and 60th percentiles respectively at the 75% satisfaction level of the requirement criteria, as depicted in figure 4.12. It should be noted that the proposed model could not find a solution for the system when the threshold values of the requirement criteria were set at the 70th percentile. This was because of the lack of availability of a sufficient number of potential stations at this threshold value (due to excessive narrowing of the search space) for generating feasible alternative solutions for the system. In this case the number of potential station locations was 1543.

These results above indicate that screening the study area for potential station locations prior to the optimization process plays an important role in directing the search towards the promising regions of the search space. Therefore, when setting the threshold values of the requirement criteria, close attention should be paid to ensure that optimal potential stations are not be excluded during this process. Based on the numerical results presented in table 6.6 it can be concluded that setting the threshold values of the requirement criteria at the 60th percentile is efficient for finding a reliable and cost effective solution for a two rail line system size. However, this conclusion may not hold for larger rail transit system sizes, i.e., rail systems with more than three lines. Therefore, it is recommended that further analysis is carried out in future to examine the sensitivity of the reduction of the number of the potential station locations on the optimality of the results. In addition, it should be noted that the satisfaction level of the requirement criteria was set at 75% for this numerical analysis. This value was set based on the sensitivity analysis carried out in section 4.4.2, which showed that setting satisfaction level at 100% resulted in an excessive reduction in the number of potential station locations due to mutual contradictions inherent in some of the requirement criteria which make their concurrent satisfaction difficult.

6.6 Impact of the Genetic Algorithm Parameters

This section tunes/refines, verifies and compares the different parameters of the genetic algorithm to find reasonable values for the parameters for solving the rail transit system planning problem.

These parameters involve population size, crossover method, crossover rate, mutation method and mutation rate. A sensitivity analysis for each of these parameters is carried out to investigate its impact on the performance of the optimization algorithm and to draw conclusions on the best settings to solve the problem. It should be noted that a solution involving a network with two rail lines, once again, the two line solution presented in section 6.2.1, is used for these analysis. This is for two main reasons: (1) using each of the possible real world solutions presented in section 6.2 (i.e., single line, two lines and three lines rail system solutions) for the sensitivity analysis would be a difficult task requiring considerable computation time and; (2) the proposed optimization algorithm applies the same concept in terms of genetic structure to represent the different line network sizes. Therefore, using a two line rail system size to perform the sensitivity analysis is sufficient to understand the impact of each parameter on the optimization algorithm search behaviour. Furthermore, since GAs are probabilistic solution algorithms, and the results are dependent on a random population, the evaluations and judgments were made based on the results obtained from three test runs. It should be noted that during the testing of each individual parameter the other remaining parameters were assumed to be constant. These constant values were determined based on preliminary tests that were carried out prior to the sensitivity analysis and were thought to be close to the values that would make the performance of the algorithm optimal. These investigations are detailed in the following subsections.

6.6.1 Impact of the Population Size

The impact of population size on the performance of the optimization algorithm is investigated in this section by applying the four test scenarios presented in table 6.7.

Table 6.7: Test Parameters for the Population Size Sensitivity Analysis

Scenarios	Population Size	No. of Generations	Crossover Method	Crossover Rate	Mutation Method	Mutation Rate
1	50	50	UCGBL	0.6	MGBL	0.1
2	100	50	UCGBL	0.6	MGBL	0.1
3	150	50	UCGBL	0.6	MGBL	0.1
4	200	50	UCGBL	0.6	MGBL	0.1

The optimized total cost for the different population sizes is shown in figure 6.5. The results show that the algorithm can find a better solution with a larger population size. The total cost of the system decreased from -100.82 M£ to -116.55 M£ (about 16%) when population size increased from 50 to 200. The results also reveal that further increment of the population size beyond 150 tends to be less efficient in terms of improving the solution. These results also indicate that with a small population size there is a risk that the algorithm would yield premature convergence due to a lack of genetic diversity in the population, thereby causing the optimization algorithm to be trapped in local optima. On the other hand, with a very large population size the algorithm tends to experience significant increases in computational time with no significant improvement in the solution fitness, as well as increased computer memory needs, which can be a problem for planning large-scale rail systems. Therefore, it is therefore important to use a population size that will allow the algorithm to converge in a reasonable time. Based on the results shown, it can be concluded that a population size of 150 is a reasonable value to find a robust solution in a reasonable computation time.

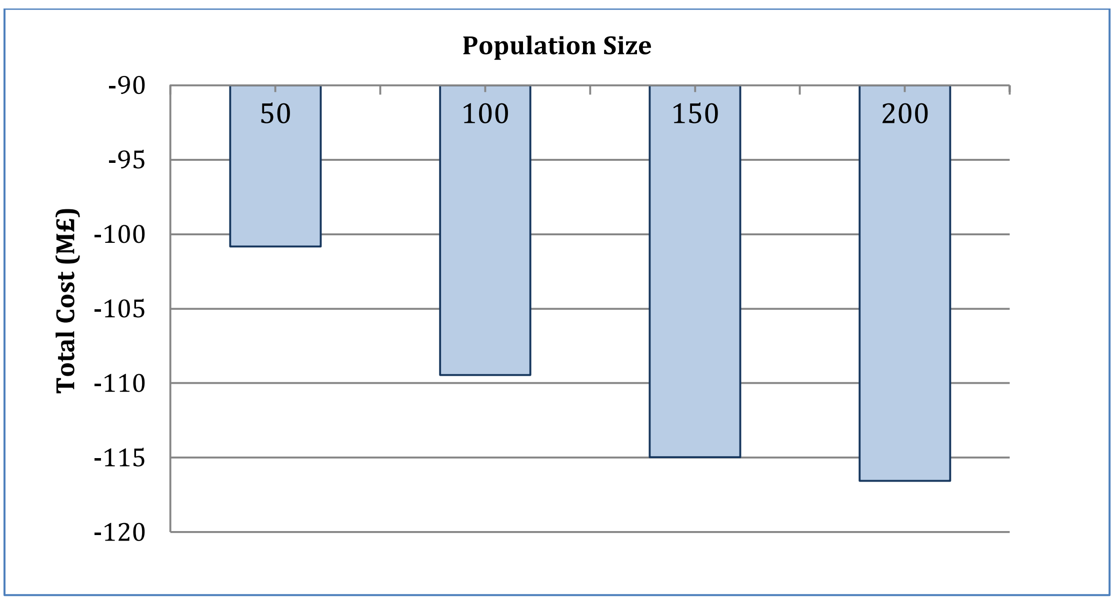


Figure 6.5: Optimized Total System Cost with Different Population Size

6.6.2 Impact of the Crossover Operator

This section examines the effect of the different crossover methods proposed in section 5.5.2 on the behaviour of the optimization algorithm in respect to directing the search towards the optimal solution and to identify the crossover method that best fits the framework of the problem solution. It also tunes the value of the crossover rate (Pc) to make the performance of the optimization algorithm optimal for solving the problem.

6.6.2.1 Crossover Method Sensitivity Analysis

The sensitivity analysis of this section compares the five crossover methods proposed in section 5.5.2 to examine the efficiency of each in terms of improving the performance of the optimization algorithm and determining the method that best fits the framework of the problem solution. This is done by applying five test scenarios, the parameters of which are shown in Table 6.8.

Table 6.8: Test Parameters for the Different Crossover Method Sensitivity Analysis

Scenarios	Population Size	No. of Generations	Crossover Method	Crossover Rate	Mutation Method	Mutation Rate
1	150	50	UCCL	0.6	MGBL	0.1
2	150	50	UCGL	0.6	MGBL	0.1
3	150	50	UCGBL	0.6	MGBL	0.1
4	150	50	OPCCL	0.6	MGBL	0.1
5	150	50	OPCGL	0.6	MGBL	0.1

The optimized total system cost with different crossover methods is shown in figure 6.6. In general, the results show that a relatively good solution is obtained from each crossover method. They also reveal that the uniform crossover at gene bit level (UCGBL) method tends to be more efficient in improving the solutions compared to the other methods. Furthermore, the optimization algorithm not only produces the best solution, but also shows a more stable search performance with this method than with the other ones. The reason why UCGBL outperforms the other methods is that it combines parent chromosomes at the gene bit (station) level while the other methods combine parent chromosomes either at gen (individual line) level or chromosome (line network) level. That is, UCGBL exchanges the corresponding stations of each two alternative solutions picked up for crossover individually, as explained in section 5.5.2.3. This reduces the probability that the new solutions would violate the required constraints and therefore, it creates a wider search space for the optimization algorithm to explore, and thereby prevents the algorithm from falling into local optima. The other crossover methods, exchange the corresponding stations of the two alternative solutions selected for crossover simultaneously either at a single line level or at the entire line network level, as detailed in sections 5.5.2.1, 5.5.2.2, 5.5.2.4 and 5.5.2.5.

Due to the complex configuration of the alternative solutions, and the need to satisfy a number of constraints, exchanging a number of stations at a time is likely to result in the production of infeasible solutions in many cases. This can be observed clearly in table 6.9, which presents the success rate of each crossover method applied in the above five test scenarios. As depicted in this table, the success rate of the UCGBL method in producing new feasible solutions is higher than the other methods. Moreover, as the level at which the stations of the alternative solutions were exchanged became coarser the success rate in producing new feasible solutions decreased. For example, since both UCGL and OPCGL operate at the individual line level, they have higher success rates (59.11% and 56.14%, respectively) than either UCCL or OPCCL (45.88% and 49.46% respectively), which operate at the line network level. The low success rate in producing new feasible solutions creates an obvious risk of losing the diversity of the search space, subsequently leading the algorithm to become trapped in local optima. Thus, UCGBL can be considered to be the most efficient crossover method among the proposed ones for rail transit system planning.

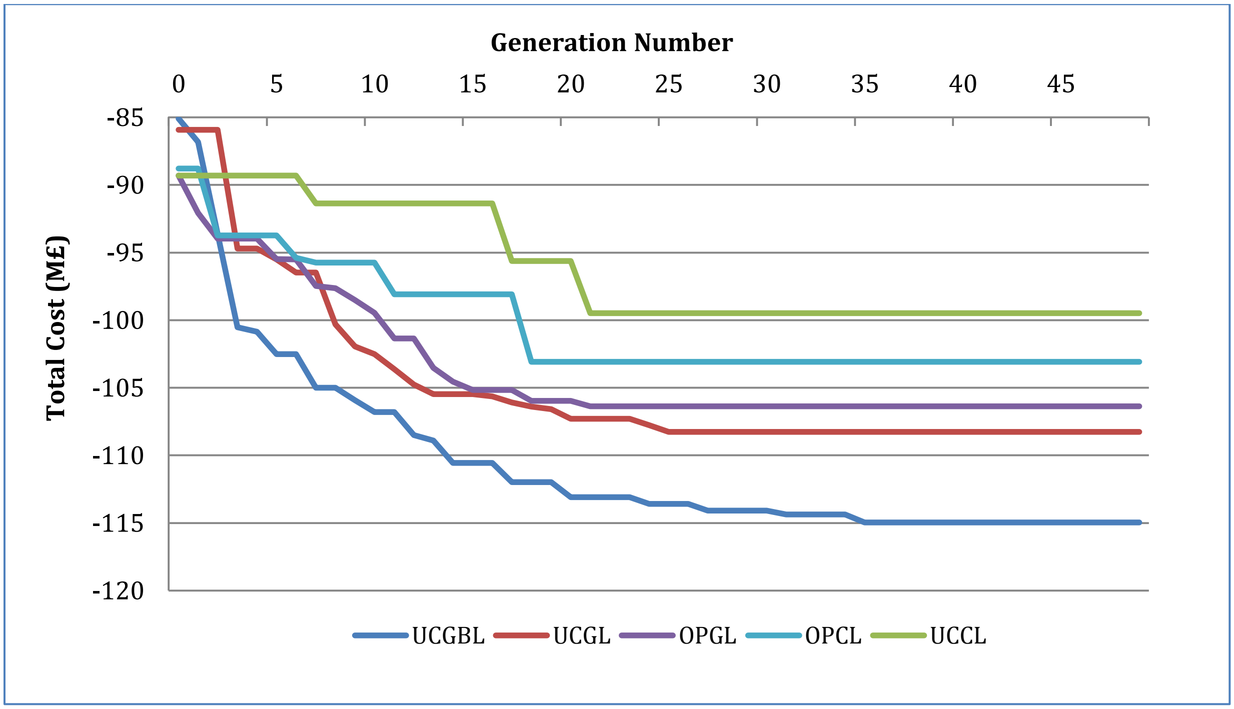


Figure 6.6: Optimized Total System Cost with Different Crossover Methods

Table 6.9: Success Rate for the Different Crossover Method

Scenarios	Crossover Method	No. of Crossover Trails	No. of the Parents Undergo Crossover (Ctf)	No. of the Generated Feasible Offspring (Ctf)	Crossover Success Rate %(Csr)
1	UCCL	2	4405	2021	45.88
2	UCGL	2	7921	4682	59.11
3	UCGBL	1	11914	8991	75.47
4	OPCCL	2	4173	2064	49.46
5	OPCGL	2	8475	4758	56.14

6.6.2.2 Crossover Rate Sensitivity Analysis

The effect of different crossover rates (Pc) on the performance of the optimization algorithm is investigated in this section by applying the six test scenarios presented in table 6.10:

Table 6.10: Test Parameters for Different Crossover Rate Sensitivity Analysis

Scenarios	Population Size	No. of Generations	Crossover Method	Crossover Rate	Mutation Method	Mutation Rate
1	150	50	UCGBL	0.5	MGBL	0.1
2	150	50	UCGBL	0.6	MGBL	0.1
3	150	50	UCGBL	0.7	MGBL	0.1
4	150	50	UCGBL	0.8	MGBL	0.1
5	150	50	UCGBL	0.9	MGBL	0.1
6	150	50	UCGBL	1	MGBL	0.1

Figure 6.7 shows the optimized total system cost with different crossover rates. The results show that the algorithm can find a better solution with a higher crossover rate, but only to a certain extent. The total cost of the system decreased by 8.72 M£ when the crossover rate increased from 0.5 to 0.7. Further increments in crossover rate beyond 0.7, however, tend to be inefficient in improving the solutions. The results also show that the best solution was obtained with a crossover rate of 0.7.

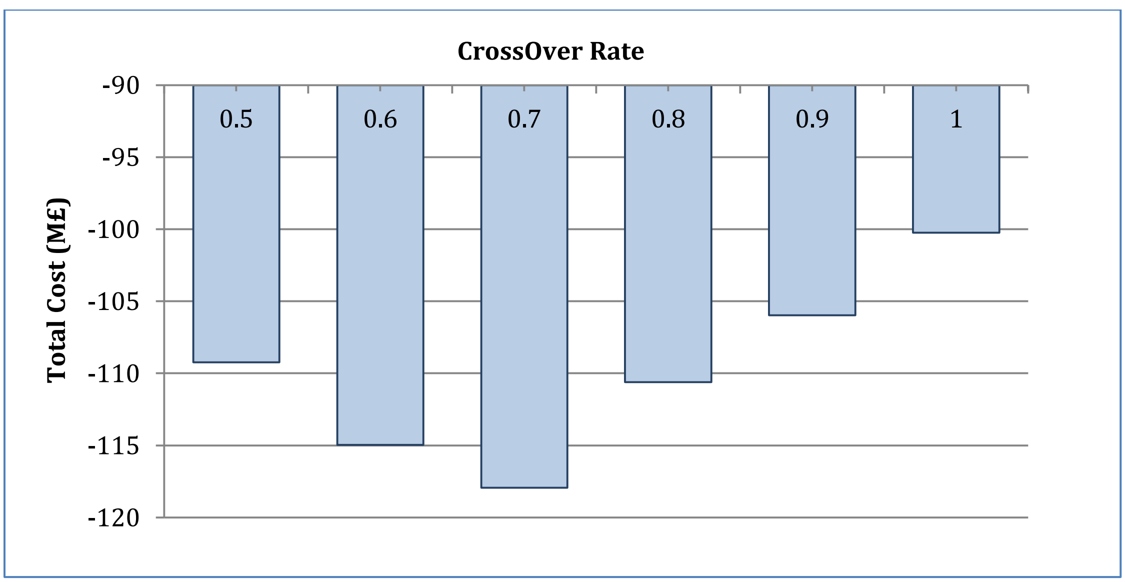


Figure 6.7: Optimized Total System Cost with Different Crossover Rates

6.6.3 Impact of the Mutation Operator

This section examines the effect of the different mutation methods proposed in section 5.5.3 on the optimization algorithm behaviour in directing the search towards the optimal solution and determining the mutation method that best fits the rail transit system planning problem. It also investigates how the performance of the optimization algorithm responds to different mutation rate (Mc) values and determines the value that can make the optimization algorithm performance optimal.

6.6.3.1 Mutation Method Sensitivity Analysis

The sensitivity analysis of this section examines the efficiency of the three mutation methods proposed in section 5.5.3 in improving the performance of the optimization algorithm and determining the method that best fits the framework of the problem solution. This is by applying the three test scenarios presented in table 6.11.

Table 6.11: Test Parameters for the Different Mutation Method Sensitivity Analysis

Scenarios	Population Size	No. of Generations	Crossover Method	Crossover Rate	Mutation Method	Mutation Rate
1	150	50	UCGBL	0.7	MCL	0.1
2	150	50	UCGBL	0.7	MGL	0.1
3	150	50	UCGBL	0.7	MGBL	0.1

The optimized total system cost with the different mutation methods is shown in figure 6.8. The results show that with the mutation at gene bit level (MGBL) method the algorithm tends to be more efficient in improving the solutions compared to the other two methods. The results also reveal that the search performance of the algorithm is more stable with this method than the others. With the MGBL method, the population improves until after the 35^{th} generation while with the MCL and MGL methods the population converges prematurely and is stuck with no further improvement beyond the 13^{th} and 27^{th} generation, respectively. These results indicate that these two methods are not able to guarantee adequate diversity of the search space. These results can be explained by the level at which these methods operate on the solutions' chromosomes. The MGBL method operates at the bit (station) level, while the MCL and MGL operate at gene (individual line) and chromosome (line network) levels respectively. That is, MGBL individually mutates a number of stations of the solution selected for mutation to produce a new solution, as explained in section 5.5.3.2. This reduces the probability that the new solution violates the required constraints, and therefore, keeps the diversity of the search space and avoids premature convergence.

The MGL and MCL methods simultaneously mutate a number of stations of the solution to produce a new solution, as explained in section 5.5.3.1and 5.5.3.2 respectively. Mutating a number of stations at a time significantly increases the probability that the new solution will be infeasible due to the complex configuration of the alternative solutions and the need to satisfy a number of constraints. This can be observed clearly in table 6.12, which presents the success rate of each mutation method applied in the above three test scenarios. The success rate of the MGBL method in producing new feasible solutions is 33.76% while the success rate of the other two methods are 0.49% and 2.76%, respectively, as depicted in table 6.12. It is also very interesting to note that the success rate of MGL is better than the success rate of the MCL method. This is simply because the MCL operates at a coarser level of the solution's chromosomes than MGL and therefore, has a higher probability of returning infeasible solutions after a mutation. Obviously, these low success rates in producing new feasible solutions result in a loss in diversity over the search space and the tendency to become stuck in local optima. Based on the results it can be concluded that the MGBL is the most efficient mutation method among those proposed for rail transit system planning.

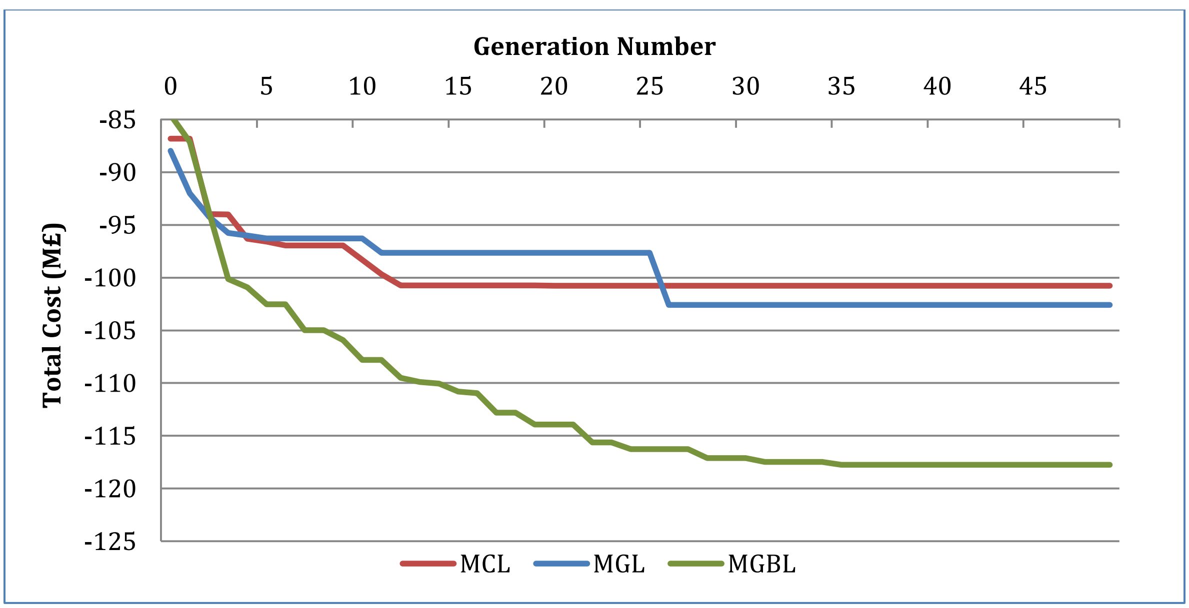


Figure 6.8: Optimized Total System Cost with Different Mutation Methods

Table 6.12: Success Rate for the Different Mutation Methods

Scenarios	Mutation Method	No. of Mutation Trails	No. of the offspring Undergo Mutation (Mtf)	No. of the Generated Feasible Offspring (Mtf)	Mutation Success Rate % (Msr)
1	MCL	2	3439	17	0.49
2	MGL	2	6047	167	2.76
3	MGBL	1	5782	1949	33.71

6.6.3.2 Mutation Rate Sensitivity Analysis

The impact of different mutation rates (Mc) on the performance of the optimization algorithm is investigated in this section by applying the five test scenarios shown in table 6.13.

Table 6.13: Test Parameters for Different Mutation Rate Sensitivity Analysis

Scenarios	Population Size	No. of Generations	Crossover Method	Crossover Rate	Mutation Method	Mutation Rate
1	150	50	UCGBL	0.7	MGBL	0.1
2	150	50	UCGBL	0.7	MGBL	0.2
3	150	50	UCGBL	0.7	MGBL	0.3
4	150	50	UCGBL	0.7	MGBL	0.4
5	150	50	UCGBL	0.7	MGBL	0.5
6	150	50	UCGBL	0.7	MGBL	0.6

The optimized total system cost with different mutation rates is demonstrated in figure 6.9. The results show that the algorithm can find better solutions with a higher mutation rate, but only to a certain extent. As depicted in figure 6.9, the total cost decreases by about 7 M£ when the mutation rate is increased from 0.1 to 0.3. However, the optimization algorithm tends to degrade the quality of the solution with further increments in the mutation rate beyond 0.3. This is because a very high mutation rate tends to lead to a significant diversification of the search space which may result in pulling good solutions away and thus preventing the algorithm from converging at any optimal solution (as discussed in section 5.5.3). The results show that the optimum solution was obtained with a mutation rate of 0.3.

Figure 6.9: Optimized Total System Cost with Different Mutation Rates

6.7 Goodness Evaluation for the Best Solution

This section aims to assess the quality of the solution found by the proposed integrated optimization model. Basically, by selecting a set of station locations and line alignments linking the stations between two terminal stations, there are a large number of possible solutions and local optima, which makes it impractical to ascertain an exact optimal solution. Since the exact optimal solution to the problem is not known (note that no existing models can guarantee finding the exact optimal solution), it is very difficult to prove the goodness of the solution found by the proposed model. This section, therefore, designs an experiment to test the goodness of the model. The experiment is initiated by finding a solution to the problem using the proposed model and then generating a set of possible solutions to the problem manually that are believed to be close to the optimal solution. Thereafter, the fitness values of the solution found by the proposed models are compared with those generated manually. The experiment considers the solution found by the model to be good if the fitness value of the solution found by the model is higher than that of all other possible solutions found manually. Using this experiment successively, the goodness of the model is tested on one of the terminal station pairs identified in section 6.2, specifically the south-north terminal station pair. Figure 6.10 illustrates the solution found by the model and some of the manually generated solutions. The reason for not presenting the whole set of the solution generated manually is just to make the figure clearer and easier to be understand.

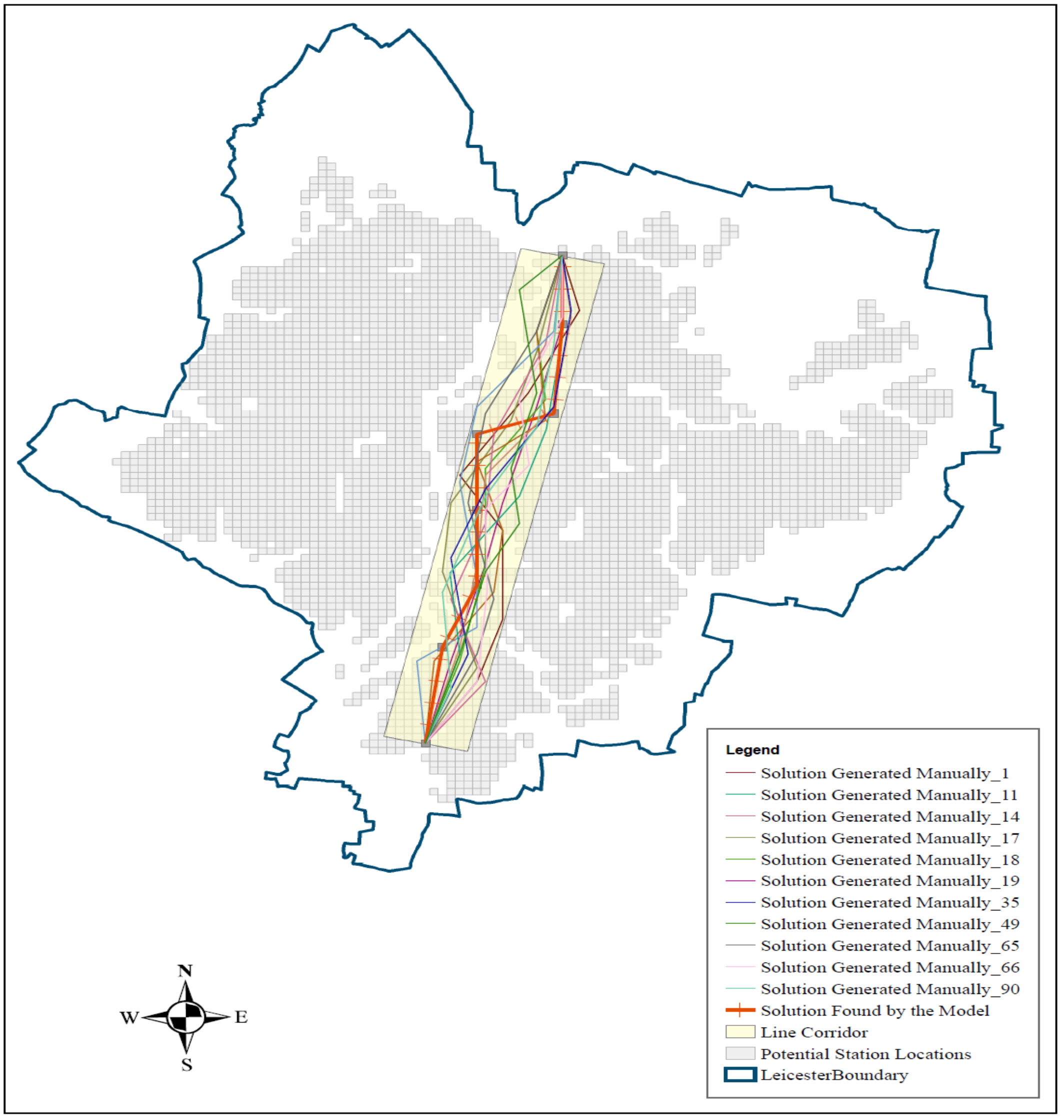


Figure 6.10: Illustration of the Proposed Model Validation

It should be noted that the width of the corridor within which the station locations and line alignment are to be determined was set at a relatively small value (1 km), as depicted in figure 6.10. This was done simply to reduce the number of possible solutions to the problem and thus to make the test more practical. Table 6.14 presents the fitness in terms of the total cost of the solution found by the proposed model and the other possible solutions found manually.

Table 6.14: Total Cost of the Solutions Found Manually and the Solutions Found by the Proposed Model

Solutions Found Manually							
Solutions No.	Total cost (M£)	Solutions No.	Total cost (M£)	Solutions No.	Total cost (M£)	Solutions No.	Total cost (M£)
1	-8.573	26	-4.972	51	-9.520	76	-3.618
2	-9.897	27	-9.416	52	-13.328	77	-8.464
3	-5.800	28	-20.670	53	-9.711	78	-25.742
4	-11.534	29	-8.259	54	-5.707	79	-13.503
5	-8.376	30	-11.444	55	-7.733	80	-8.998
6	-7.522	31	-11.242	56	-7.099	81	-19.268
7	-10.647	32	-8.763	57	-6.069	82	-8.306
8	-6.775	33	-8.152	58	-9.098	83	-18.730
9	-10.333	34	-10.442	59	-9.999	84	-21.841
10	-11.840	35	-13.031	60	-9.245	85	-29.186
11	-8.462	36	-7.874	61	-12.909	86	-8.262
12	-5.381	37	-8.625	62	-4.330	87	-5.577
13	-13.788	38	-10.831	63	-11.721	88	-11.791
14	-12.375	39	-5.239	64	-7.029	89	-12.291
15	-13.570	40	-7.160	65	-15.581	90	-17.801
16	-11.661	41	-9.485	66	-12.418	91	-10.476
17	-13.340	42	-9.029	67	-9.210	92	-7.666
18	-7.676	43	-6.055	68	-9.917	93	-6.404
19	-9.462	44	-7.217	69	-8.313	94	-9.054
20	-7.633	45	-12.490	70	-10.942	95	-9.724
21	-9.222	46	-7.230	71	-8.622	96	-11.050
22	-12.352	47	-14.689	72	-17.188	97	-8.668
23	-11.740	48	-7.176	73	-13.831	98	-4.744
24	-12.043	49	-19.969	74	-7.258	99	-11.359
25	-5.028	50	-7.439	75	-8.860	100	-9.035

Solutions Found by the Proposed Integrated Optimization Model	
Solutions No.	Total cost (M£)
1	-36.306

As depicted in table 6.14, the solution found by the proposed model possesses the highest fitness value of the other possible solutions to the problem and therefore, it can be concluded that the proposed model can find a very good solution to the problem.

6.8 Summary

This chapter presents the application of the proposed integrated optimization model to a real world case study, the City of Leicester (UK), in order to examine its effectiveness in finding a robust solution and to demonstrate its applicability in real world planning practice. An extensive numerical study was also included in this chapter to: examine the sensitivity of the proposed model with demand variation; compare individual and simultaneous optimization methods for planning a rail transit system; examine the importance of the initial identification of the potential station locations in directing the search process towards optimal solutions. This is in addition to examining different parameters and structures of the GA operators on the model performance and evaluating the goodness of the solution found by the proposed model.

The case study demonstrates the applicability of the proposed model not only for planning a new rail transit system but also for expanding an existing one. The results show that the proposed model can effectively resolve the essential trade-off between maximum rail system usage and minimum passenger travel time, on the one hand, and the minimum construction cost of the system on the other hand, while also complying with the required constraints. The results also show that performing a feasibility analysis for potential station locations prior to the optimization process can effectively improve the performance of the model to realize a robust solution by directing the search to explore the promising regions in the search space. Furthermore, the numerical results prove the advantage of simultaneous optimization of multiple rail lines over the individual line optimization in finding a cost-effective solution. The goodness evaluation of the model proves that the model can find a very good solution. It should be noted that some of the data used here had no values on the uncertainty bounds. The effect of uncertainty in the relevant aspects of the data should be analysed in future work.

The next chapter concludes the thesis by setting out the main outcomes of the research, including potential topics for future research in rail transit system planning field.

Chapter 7 – Conclusions and Future Work

This chapter presents the conclusions drawn from the work presented in this thesis, in relation to its stated aims and objectives. Recommendations for future work are also given.

7.1 Conclusions

As described in section 1.2, the aim of this thesis is to develop an optimal planning method that treats the rail network system and its influencing factors in a single integrated process in order to help rail transit planners to produce optimal (reliable and cost effective) rail transit systems. In order to realize this aim, the objectives of the research were as follows:

1- Identify the requirements for rail transit system planning with respect to the passenger level of service, operator productivity and the potential benefits to the community, each of which have a significant influence on both the location and configuration of the rail transit system.
2- Disaggregate the identified requirement sets of the three interrelated parties: passengers, operators and the community, into group sets according to their interactions with the two main components of rail transit system planning (station locations and the alignment of the line network connecting the stations).
3- Quantify and formulate various associated station and rail line network planning requirements as an optimization problem in order to achieve an efficient and effective rail transit system.
4- Develop an effective method to seek the best solution for the rail transit system planning problem with respect to the formulated set of requirements and constraints.
5- Conduct a real world case study to examine the effectiveness of the proposed method and to confirm its validity.

The work presented in this thesis meets all the objectives above. The first three objectives are met by the development of the new planning framework, presented in chapter 3. The framework brings together various planning requirements and incorporates them into a single planning platform. This is to ensure that the proposed rail transit system has a positive effect on the area it serves, including the mitigation of congestion, improvements in mobility, economic development and environmental enhancement. It consists of three main levels of analysis and decision-making. Level I identifies the requirements that must be accounted for in rail transit system planning, based on a detailed and comprehensive literature review. These requirements involve the consideration of the level of service to passengers, operator productivity and potential benefits for the community. The analysis and decision making process at level II translates these requirements into effective criteria that can be used to evaluate various alternative solutions. The translation of the requirements into sets of criteria is performed based on a comprehensive literature review of rail system planning and the factors upon which the identified requirements depend. Level III formulates mathematical functions for these criteria, and incorporates them into a single planning platform within the context of an integrated optimization model in order to achieve a rail transit system that best fits the desired requirements identified at level I.

The fourth objective is met by the development of the integrated optimization model, which simultaneously determines station locations and the line network alignment connecting the stations in two stages. The first stage is embedded within a GIS and screens the study area for a set of feasible station locations with a comprehensive consideration of the various requirements (identified at level I) based on systematic evaluation and comparative analysis. The second stage uses a heuristic optimization algorithm based on GA and supported by the background GIS database simultaneously to select the optimal set of station locations from the pool of feasible stations and to generate the line network connecting these stations. The modelling framework resolves the essential trade-off between an effective rail system that provides high service quality and benefits for both the passenger and the whole community, and an economically efficient system with acceptable capital and operational costs.

The fifth objective is met by applying the developed integrated optimization model to the real world case study of the City of Leicester in the UK. The case study demonstrates the practical applicability of the model, not only for planning a new rail system with multiple lines but also for expanding existing ones. The results reveal that the model can effectively resolve the essential trade-off between maximum rail system usage and minimum passenger travel time, on the one hand, and the minimum construction cost of the system on the other hand, while also complying with the various constraints. Furthermore, the results reveal that performing a feasibility analysis for potential station locations prior to the optimization process can effectively improve the performance of the model to realize an optimal (reliable and cost effective) solution, by directing the search to explore the most promising regions in the search space. The total cost of the system decreases by 27.56% when the threshold values of the requirement criteria are increased from the 30th to the 60th percentiles for screening the study area for feasible station locations. When setting the threshold values of the requirement criteria, however, attention should be paid to ensure that optimal potential stations are not excluded during this process. Based on the numerical results, it is concluded that setting the threshold values of the requirement criteria at the 60th percentile is efficient for finding a reliable and cost effective solution for a two rail line system size. Furthermore, the numerical results confirm the advantage of the simultaneous optimization of multiple rail lines compared to individual line optimization in finding a cost-effective solution. Simultaneous optimization reduced the total cost by about 70% in respect to individual line optimization. The goodness evaluation of the model proves that the model can find a very good solution.

In summary, this thesis has made the following main findings/contributions:

1- The limitations of the existing empirical work and theoretical studies for rail transit system planning have been identified and a new method proposed. This method can automatically generate solution alternatives in a very efficient manner, both for planning a new rail transit system and expanding existing ones, while considering various local conditions and the multiple requirements that arise from passengers, operators and the community.
2- A new framework has been developed that brings together the various planning requirements of the rail transit system's different stakeholders (passengers, operators and the community), and integrates them into a single planning platform. This is to ensure that the proposed rail transit system brings positive impacts to the area it serves, including the mitigation of congestion, improvements in mobility, economic development and environmental enhancement.
3- An integrated optimization model has been developed that integrates complex correlations and interactions between rail transit station locations and line alignments considering multiple rail lines which is largely neglected in the existing literature. The model formulations address the essential trade-off between maximum rail system usage and minimum passenger travel time, on

the one hand, and the minimum construction cost of the system on the other hand, while also complying with various constraints.

4- A comprehensive cost evaluation function has been developed for evaluating and comparing different alternative solutions and this has been integrated with background GIS database. This function accounts for various local conditions of the study area including land use pattern, rights-of-way, topography and geology, which are largely ignored in the existing literature.

5- Tunnel cost estimation models have been developed that can be used for various applications in the planning stage based on statistical analysis of historical cost data while taking into account tunnel size and length, geological conditions and excavation methods. In addition, a heuristic algorithm has been developed to determine the optimal tunnel depth along a rail transit line network, considering intersections of lines and integrating this into the tunnel cost estimate model to compute the total construction cost of the tunnel structures. The impact of tunnel depth and ground condition through which tunnels are bored on the tunnel estimation cost is ignored in the existing literature by assuming that the tunnel cost is captured by the linear function of tunnel length and diameter.

6- Different specific genetic operators (crossover and mutation) have been designed for evolving alternative solutions towards the optimal solution, evaluating their goodness in facilitating the efficiency of the search process and directing that search process towards the promising regions in the search space. Ultimately, the structure and parameters of these operators that best fits the framework of rail transit system planning are found.

7.2 The Implications of the Thesis

This thesis has developed an integrated optimization model that can be used efficiently both for planning a new rail transit system and for expanding an existing one while taking into account the various local conditions and multiple requirements that arise from the different stakeholders in rail transit system planning; passengers, operators and the community. The City of Leicester is selected in this thesis as a case study to examine the effectiveness of the model in finding a solution that best fits the desired objectives. However, the model is designed in such a way that it can be applied to any city in the world. The results show that the model is able to generate “good” solutions even in areas with complex geographical features, as illustrated in chapters 4 and 6. Additional practical applications of the model can be inferred based on these results, as follows:

1- The model can automatically generate alternative solutions in a very efficient manner, which significantly reduces the time required to plan and set out a rail transit system project and therefore reduces the total cost of the entire process.
2- The model not only considers the complex correlations and interactions between rail transit station locations and line alignments by integrating these two intertwined components in a single optimization process, but also considers simultaneous optimization of multiple rail lines. This reduces the total system cost significantly, and thus achieves a cost effective solution. For example, the analysis results showed that simultaneous optimization of multiple rail lines reduced the total system cost by about 70% in respect to individual line optimization, which is what is largely applied in current real world practice.
3- The model allows for comprehensive coordination with patterns of land use, particularly high-density commercial land uses (like central business districts (CBD), recreational centres, and office complexes), and existing transport networks, while seeking for the solutions, in particular in solutions for station locations. It therefore contributes to promoting sustainable development in the city/area under consideration by providing efficient mobility and accessibility to activities surrounding the system, boosting business activity and economic productivity, as well as coaxing people out of their cars and into trains. Such developments are very important for increasing the acceptance rate of the project among the public and decision makers.
4- The model incorporates the flexibility to respond to the different requirements of the rail transit system planners and policy makers. This is achieved by embedding weighting factors into the fitness function (see equation 5.1); i.e. a higher or lower weight can be given to the requirement set of each particular rail transit system stakeholder thereby prioritizing or neglecting the desirability of that particular requirement set. This also allows the rail transit system planners and policy makers to examine the impact of the requirements of each particular stakeholder on the final configuration of the proposed rail system and its effectiveness in achieving the desired objectives.

5- The model incorporates a comprehensive cost evaluation function, which is represented in terms of passenger, operator and the community costs and is embedded into the fitness function to evaluate and compare different alternative solutions. This function is designed in such a way that it incorporates the ability to be used effectively outside the model framework, thereby promoting fuller insights into the planning problem on the part of the different stakeholders. For example, the local government or the public sector, which supplies the capital costs for constructing the proposed rail transit system, can use the community cost to estimate the total construction cost of the system in the first instance. This is very important to determine whether this cost can be afforded or whether there is a need to involve other public sectors to share this construction cost, as well as to assess whether this investment cost is worth the potential benefits of the system. Furthermore, it can be used to examine the impact of the various parameters, such as tunnelling methods, geological conditions, topography and right of way costs on the system total construction cost. In addition, the transport network operator, who is responsible for managing, operating and maintaining the proposed rail system, can utilize the operator cost function to estimate the potential revenue that can be achieved from the system throughout its life span. Moreover, this cost function can be utilized to examine how various operational parameters, such as station density, operating speed and line alignment length, influence the system operator productivity level.

7.3 Recommendations for Future Work

This thesis recommends the following future enhancement of the proposed optimization model.

- The proposed model, particularly its cost evaluation framework, is designed in such a way that it incorporates the flexibility to be expanded to further enhance the practical applicability of the proposed model. The current cost evaluation function includes rights-of-way costs, station building costs, track costs, tunnel costs, escalator barrel costs, passenger travel time costs and operation and maintenance costs. This cost evaluation function, which is tailored for planning underground/metros rail transit system, can be modified for the model to be applicable for above ground rail transit systems. This can be done by including bridge construction costs and earthwork costs.

- The proposed model considers some general geometric requirements for rail line alignments such minimum and maximum gradient but does not account for the smoothness of the line alignment, i.e., geometric requirements for horizontal and vertical line alignments. It is recommended that more specific geometric constraints for line alignments such as minimum horizontal curve radius, minimum vertical curve radius, minimum tangent length and minimum spiral rate are incorporated within the optimization framework.
- More sensitivity analysis of the effects of satisfaction level and threshold values on the performance of the proposed model in finding good solutions for different sizes of rail transit systems can be performed to make the model more robust and applicable.
- Future work should address the effect of input data quality.

References

Babalik, E. 2000. "*Urban Rail Systems: A Planning Framework to Increase Their Success*", Ph.D. Thesis, University of London.

Bertolini, L. 1999, "Spatial Development Patterns and Public Transport: The Application of an Analytical Model in the Netherlands", *Planning Practice & Research*, Vol.14, pp. 199-210.

Bertaud, A. 2002, "Note on Transportation and Urban Spatial Structure". In Proceeding of Annual Bank Conference on Development Economics, Washington, D.C.

Blickle , T. and Thiele, L., 1995, "A Comparison Of Selection Schemes Used In Genetic Algorithms",TIK-Report,

Bowman, L. A. and Turnquist, M. A. 1981, "Service Frequency, Schedule Reliability and Passenger Wait times at Transit Stops", *Transportation Research Part A: General*, Vol.15, n. 6, pp. 465-471.

Brons, M., Givoni , M. and Rietveld, P. 2009, "Access to Railway Stations and Its Potential in Increasing Rail Use", *Transportation Research Part A*, Vol.43, pp. 136–149.

Bruno, G. and Laporte, G. 2002, "An Interactive Decision Support System for the Design Of Rapid Public Transit Networks", *Information Systems and Operational Research*, Vol.40, n. 2, pp. 111-118.

Bruno, G., Gendreau, M. and Laporte, G. 2002, "A Heuristic for The Location of A Rapid Transit Line", *Computers & Operations Research*, Vol.29, n. 1, pp. 1-12.

Carrizosa, E., Harbering, J. and Schöbel, A. 2016, "Minimizing the passengers' traveling time in the stop location problem", *Journal of the Operational Research Society, DOI:10.1057/jors.2016.3*, pp. 1-13.

Cervero, R., 1993, "Ridership Impacts of Transit-Focused Development in California",Working Paper,University of California Transportation Center.

Cervero, R. 2001, "Walk-And-Ride: Factors Influencing Pedestrian Access To Transit", *Journal of Public Transportation*, Vol.7, n. 3, pp. 1-23.

Cervero, R. and Guerra, E. 2011, "Urban Densities and Transit: A Multi-dimensional Perspective", *Working Paper, Institute Of Transportation Studies , University Of California, Berkeley*.

Cervero, R. and Kockelman, K. 1997, "Travel Demand and the 3Ds: Density, Diversity, and Design", *Transportation Research Part D: Transport and Environment* Vol.2, n. 3, pp. 199-219.

Cervero, R. and Murakami, J., 2008, "Rail + Property Development: a Model of Sustainable Transit Finance and Urbanism",Working Paper,Berkeley: UC Berkeley Center for Future Urban Transport: A Volvo Center of Excellence.

Cervero, R., Murphy, S., Ferrell, C., Goguts, N., Tsai, Y.-H., Arrington, G. B., Boroski, J., Smith-Heimer, J., Golem, R., Peninger, P., Nakajima, E., Chui, E., Dunphy, R., Myers, M., Mckay, S. and Witenstein, N., 2004, "Transit-Oriented Development in the United States: Experiences, Challenges, and Propspects",TCRP Report 102, Transportation Research Board,Washington, D.C.

Chapleau, R., Lavigueur, P. and Baass, K. 1987. A Posteriori Impact Analysis of a Subway Extension in Montreal. *Publication CRT-503.* Centre de recherche sur les transports, Montreal, Canada.

Chatman, D. G., Noland, R. B., Tulach, N. K., Grady, B. P., Ozbay, K., Rognlien, L., Desautels, A., Alexander, L. and Sgleave, S. D., 2012, "*Methodology for Determining the Economic Development Impacts of Transit Projects"*. http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_w56.pdf.

Chen, C., Gong, H. and Paaswell, R. 2008, "Role of the Built Environment on Mode Choice Decisions: Additional Evidence on the Impact of Density", *Transportation*, Vol.35, pp. 285–299.

Chen, J.-C. 2003, "Dijkstra's Shortest Path Algorithm", *Journal of Formalized Mathematics*, Vol.15, pp. 1-9.

Chester, M. and Horvath, A., 2008, "*Environmental Life-cycle Assessment of Passenger Transportation: A Detailed Methodology for Energy, Greenhouse Gas and Criteria Pollutant Inventories of Automobiles, Buses, Light Rail, Heavy Rail and Air v.2"*. http://escholarship.org/uc/item/5670921q.

Chien, S. and Schonfeld, P. 1997, "Optimization of Grid Transit System In heterogeneous Urban Environment", *Journal of Transportation Engineering*, Vol.123, pp. 28-35.

Coley, D. A. 1999, *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific Publishing Co. Pte.Ltd., Singapore.

Costa, A. L., Cunha, C. and Coelho, P. a. L. F. 2016, "Decision Support Systems for Real-World High-Speed Rail Planning", *Journal of Transportation Engineering*, Vol.142, n. 5, pp. 1-10.

Costa, A. L., Cunha, C., Coelho, P. a. L. F. and Einstein, H. H. 2013, "Solving High-Speed Rail Planning With The Simulated Annealing Algorithm", *Journal of Transportation Engineering*, Vol.139, n. 6, pp. 635-642.

Cts, 2009, "*Understanding the Impacts of Transitways: The Hiawatha Line: Impacts on Land Use and Residential Housing Value*", Center for Transportation Studies, University of Minnesota. http://www.cts.umn.edu/Research/Featured/Transitways at www.cts.umn.edu/Publications/ResearchReports/pdfdownload.pl?id=1222.

Debrezion, G., Pels, E. and Rietveld, P. 2007, "The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis", *Journal of Real Estate Finance and Economics*, Vol.35, pp. 161–180.

De Rus, G., 2012, "*Economic evaluation of the High Speed Rail*", Ministry of Finance, Sweden. Available at: http://www.ems.expertgrupp.se/uploads/documents/hsr.pdf Accessed 19 February 2016.

Dufourd, H., Gendreau, M. and Laporte, G. 1996, "Locating A Transit Line Using Tabu Search", *Transportation Science*, Vol.4, n. 1–2, pp. 1-19.

Evans, I., Pratt, R. H., Stryker, A. and Kuzmyak , J. R. 2007 "Chapter 17—Transit Oriented Development", TCRP Report 95: Traveler Response toTransportation System Changes. Transportation Research Board, Washington, D.C.

Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L. 2004, "How (In)accurate are Demand Forecasts in Public Works Projects: the Case of Transportation", *Journal of the American Planning Association*, Vol.71, pp. 131-146.

Frank , L. D. and Pivo, G. 1994, "Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking", *Transportation Research Record* n. 1466, pp. 44-52.

Garcia, A., 2010, "Relationship Between Rail Service Operating Direct Costs And Speed",International Union of Railways,Paris.

Gard, J. 2007, "Innovative Intermodal Solutions for Urban Transportation Paper Award: Quantifying Transit-Oriented Development's Ability To Change Travel Behavior", *ITE Journal*, Vol.77, n. 11, pp. 42-46.

Garrett, T. A. and Castelazo, M. D., 2004, "Light Rail Transit in America: Policy Issues and Prospects for Economic Development",Report,Federal Reserve Bank of St. Louis.

Givoni, M. and Rietveld, P. 2007, "The Access Journey to the Railway Station and its Role in Passengers' Satisfaction with Rail Travel", *Transport Policy* Vol.14, pp. 357–365.

Gleave, S. D. 2005, "What Light Rail Can Do for Cities: A Review of the Evidence", Research Report, Passenger Transport Executive Group, UK.

Goldberg, D. E. 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Massachusetts.

Goldberg , D. E. and Deb, K. 1991, "A Comparison Analysis Of Selection Schemes Used In Genetic Algorithms", In: G.J. E. Rawlins (ed.),Foundations of Genetic Algorithms. Morgan Kaufmann, San Mateo, California, pp. 69-93.

Goverde, R. M. P. 2005. "*Punctuality of Railway Operations and Timetable Stability Analysis*", Ph.D. Dissertation, Delft University of Technology.

Gutiérrez-Jarpa , G., Obreque , C., Laporte , G. and Marianov, V. 2013, "Rapid Transit Network Design For Optimal Cost And Origin–Destination Demand Capture", *Computers & Operations Research*, Vol.40, pp. 3000-3009.

Hamacher, H. W., Horn, S. and Schöbel, A. 2009, "Stop Location Design In Public Transportation Networks: Covering And Accessibility Objectives", *TOP*, Vol.17, n. 2, pp. 335-346.

Hay, W. W. 1982, *Rail Road Engineering*, John Wiley & Sons, New York.

Hazledine, T., Donovan, S. and Bolland, J., 2013, "*The Contribution Of Public Transport To Economic Productivity*", Nz Transport Agency Research Report 514. http://www.nzta.govt.nz/resources/research/reports/514/docs/514.pdf.

Hong, J., Shen, Q. and Zhang, L. 2014, "How Do Built-Environment Factors Affect Travel Behavior? A Spatial Analysis at Different Geographic Scales", *Transportation*, Vol.41, pp. 419-440.

Howarth, H., Griffin, M., Childs, C., Fujiyama , T., Havenith, G. and Hodder, S., 2011, "Comfortable Sardine:The balance between comfort and capacity",UK.

Jha, M. K. 2000. "*A Geographic Information Systems-Based Model for Highway Design Optimization*", Ph.D. Dissertation, University of Maryland.

Jhk and Associates, 1987, "Development-Related Survey I Washington, D.C.",Washington Metropolitan Area Transit Authority.

Jha, M. K. and Kim, E. 2006, "Highway Route Optimization Based on Accessibility, Proximity, and Land-Use Changes", *Journal of Transportation Engineering*, Vol.132, pp. 435.

Jha, M. K. and Oluokun, C. 2004, "Optimizing Station Locations Along Transit Rail Lines with Geographic Information Systems and Artificial Intelligence", In: Allan, J., Brebbia, C.A., Hill, R.J., Sciutto, G., Sone, S. (Eds.), Computers in Railways IX (COMPRAIL 2004). WIT Press, Southampton, UK.

Jha, M. K. and Schonfeld, P. 2004, "A Highway Alignment Optimization Model Using Geographic Information Systems", *Transportation Research Part A*, Vol.38, n. 6, pp. 455-481.

Jha, M. K., Schonfeld, P. and Kim, E. 2006, *Intelligent Road Design*, WIT Press, Southhampton, UK.Jha, M. K. and Schonfeld, P. 2007, "Optimizing Rail Transit Routes with Genetic Algorithms and Geographic Information System", *Journal of Urban Planning and Development*, Vol.133, pp. 161-171.

Jha, M. K., Schonfeld, P. and Samanta, S. 2007, "Optimizing Rail Transit Routes with Genetic Algorithms and Geographic Information System", *Journal of Urban Planning and Development*, Vol.133, n. 3, pp. 161-171.

Johnson, D. B. 1973, "A Note on Dijkstra's Shortest Path Algorithm", *Journal of the ACM (JACM)*, Vol.20, n. 3, pp. 385-388

Jong, J. C. 1998. "*Optimizing Highway Alignments with Genetic Algorithms*", Ph.D. Dissertation, University of Maryland, College Park.

Jong, J. C. and Schonfeld, P. 2003, "An Evolutionary Model For Simultaneously Optimizing Three-Dimensional Highway Alignments", *Transportation Research Part B*, Vol.37, n. 2, pp. 107-128.

Kamruzzaman , M., Baker, D., Washington , S. and Turrell , G. 2014, "Advance Transit Oriented Development Typology: Case Study in Brisbane, Australia", *Journal of Transport Geography*, Vol.34, pp. 54-70.

Kang, M. W. 2008. "*An Alignment Optimization Model For a Simple Highway Network*", Ph.D. Dissertation, University of Maryland, College Park.

Kang, M. W., Jha, M. K. and Buddharaju, R. 2014, "Rail Transit Route Optimization Model for Rail Infrastructure Planning and Design: Case Study of Saint Andrews, Scotland", *Journal of Transportation Engineering*, Vol.140, n. 1, pp. 1-11.

Kardas-Cinal, E. 2010, "Ride comfort for various passenger positions in a railway vehicle simulation study", *Archives of Transport*, Vol.22, n. 2, pp. 182-200.

Kaul, K. 2010, *Cut-and-Cover Metro Structures Geo-Structural design: An Integrated Approach*, Spon Press, New York, USA.

Keijer, M. J. N. and Rietveld, P. 2000, "How Do People Get To The Railway Station? The Dutch Experience", *Transportation Planning and Technology*, Vol.23, pp. 215-235.

Kikuchi, S. and Vuchic, V. R. 1982, "Transit Vehicle Stopping Regimes and Spacings", *Transportation Science*, Vol.16, n. 3, pp. 311-331.

Kim, E. 2001. "*Modeling Intersections & Other Structures in Highway Alignment Optimization*", Ph.D. Dissertation, University of Maryland, College Park.

Kim, E., Jha, M. K. and Son, B. 2005, "Improving The Computational Efficiency of Highway Alignment Optimization Models Through a Stepwise Genetic Algorithms Approach", *Transportation Research Part B*, Vol.39, n. 4, pp. 339-360.

Kim, S., Ulfarsson, G. F. and Hennessy, T. J. 2007, "Analysis of Light Rail Rider Travel Behavior: Impacts of Individual, Built Environment, and Crime Characteristics on Transit Access", *Transportation Research Part A* Vol.41, pp. 511–522.

Kitamura, R., Mokhtarian, P. L. and Laidet, L. 1997, "A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area", *Transportation*, Vol.24, pp. 125-158.

Knowles, R. D. 1996, "Transport impacts of Greater Manchester's Metrolink Light Rail System", *Journal of Transport Geography*, Vol.4, n. 1, pp. 1-14.

Knutm, D. E. 1977, "A generalization of Dijkstra's Algorithm", *Information Processing Letter*, Vol.6, n. 1, pp. 1-5.

Koppelman, F. S. and Bhat, C. 2006, *A self instructing course in mode choice modeling: multinomial and nested logit models*, FTA.

Lai, X. 2012. "*Optimization of Station Locations and Track Alignments for Rail Transit Lines*", Ph.D. Dissertation, University of Maryland, College Park.

Lai, X. and Schonfeld, P. 2010, "Optimizing Rail Transit Alignments Connecting Several Major Stations". In Proceeding of Compendium of Papers of the 89th TRB Annual Meeting, Washington, D.C.

Lai, X. and Schonfeld, P. 2012, "Optimization of Rail Transit Alignments Considering Vehicle Dynamics", *Transportation Research Record,Journal of the Transportation Research Board*, n. 2275, pp. 77-87.

Lai, X. and Schonfeld, P. 2016, "Concurrent Optimization of Rail Transit Alignments and Station Locations", *Urban Rail Transit*, Vol.2, pp. 1-15.

Laporte, G., Marín, Á., Mesa, J., and Ortega, F. 2007. "*An Integrated Methodology for the Rapid Transit Network Design Problem*". *In:* Geraets, F., Kroon, L., Schoebel, A., Wagner, D. & Zarioliagis, C. (eds.), *Algorithmic Methods for Railway Optimization. Lectures Notes in Computer Science.* Springer, Berlin, v.4359**,** pp.187-199.

Laporte, G., Mesa, J. A. and Ortega, F. A. 2000, "Optimization Methods for the Planning of Rapid Transit Systems", *European Journal of Operational Research*, Vol.122, n. 1, pp. 1-10.

Laporte, G., Mesa, J. A. and Ortega, F. A. 2002, "Locating Stations on Rapid Transit Lines", *Computers & Operations Research*, Vol.29, n. 6, pp. 741-759.

Laporte, G., Mesa, J., Ortega, F. and Sevillano, I. 2005, "Maximizing Trip Coverage in the Location of a Single Rapid Transit Alignment", *Annals of Operations Research*, Vol.136, n. 1, pp. 49-63.

Laporte, G., Mesa, J. A., Ortega, F. A. and Perea, F. 2011, "Planning rapid transit networks", *Socio-Economic Planning Sciences*, Vol.45, n. 3, pp. 95-104.

Lin, J. and Shin, T. 2008, "Does Transit-Oriented Development Affect Metro Ridership?: Evidence from Taipei, Taiwan ", *Journal of the Transportation Research Board*, Vol.2063, pp. 49–158.

Litman, T., 2010, "*Evaluating Transportation Economic Development Impacts"*, Victoria Transport Policy Institute Transportation. http://vtpi.org/econ_dev.pdf.

Litman, T. 2012, "Rail Transit in America: A Comprehensive Evaluation of Benefits", Research Report, VictoriaTransport Policy Institute, Canada.

Litman, T., 2015a, "*Evaluating Public Transit Transit As An Energy Conservation and Emission Reduction Strategy ",* Victoria Transport Policy Institute. http://www.vtpi.org/tran_climate.pdf.

Litman, T., 2015b, "*Evaluating Public Transit Benefits and Costs"*, Victoria Transport Policy Institute. www.vtpi.org/tranben.pdf.

Loo, B. P. Y., Chen, C. and Chan, E. T. H. 2010, "Rail-Based Transit-Oriented Development: Lessons from New York City and Hong Kong", *Landscape and Urban Planning*, Vol.97, pp. 202-212.

Loutzenheiser, D. 1997, "Pedestrian Access to Transit: Model of Walk Trips and Their Design and Urban Form Determinants Around Bay Area Rapid Transit Stations ", *Transportation Research Record 1604*, pp. 40-49.

Mackett, R. L. and Edwards, M. 1998, "The Impact of New Urban Public Transport Systems: Will the Expectations be Met?", *Transportation Research A*, Vol.32 n. 4, pp. 231–245.

Marín, Á. 2007, "An Extension to Rapid Transit Network Design Problem", *Top*, Vol.15, n. 2, pp. 231-241.

Mohan, D., 2005, "*Public Transportation Systems for Urban Areas A Brief Review"*, Transportation Research and Injury Prevention Programme and Indian Institute of Technology Delhi. Available at: http://tripp.iitd.ernet.in/delhibrts/metro/Metro/public%20trpt%20systm%20-%20review.pdf Accessed 1 september 2016.

O'Sullivan, S. and Morrall, J. 1996, "Walking Distances To And From Light-Rail Transit Stations", *Transportation Research Record 1538*, pp. 19–26.

Olssona, N. O. E. and Haugland, H. 2004, "Influencing Factors On Train Punctuality—Results From Some Norwegian Studies", *Transport Policy*, Vol.11, pp. 387–397.

ORR, 2012, "*Costs and Revenues of Franchised Passenger Train Operators in the UK*", Office of Rail Regulation, UK. Available at: http://orr.gov.uk/__data/assets/pdf_file/0019/4933/toc-benchmarking-report-2012.pdf Accessed 13 March 2016.

Parker, N. A. 1977, "Rural Highway Route Corridor Selection", *Transportation Planning and Technology*, Vol.3, pp. 247-256.

Patil, V. P. and Pawar, D. D. 2015, "The Optimal Crossover Or Mutation Rates In Genetic Algorithm: A Review", *International Journal of Applied Engineering and Technology*, Vol.5, n. 3, pp. 38-41.

Piszcz , A. and Soule, T. 2006, "Genetic Programming: Analysis of Optimal Mutation Rates in a Problem with Varying Difficulty", *International Journal of Innovative Computing and Applications*, Vol.1, n. 2, pp. 108-120.

PR103 Demand Model Development Report, 2013, "*Leicestershire County Council*", UK. Available at: http://www.leics.gov.uk/llitm.htm Accessed 19 September 2015.

Reilly, M. and Landis, J., 2002, "The Influence of Built-Form and Land Use on Mode Choice. Evidence from the 1996 Bay Area Travel Survey",University of California Transportation Center Research Paper, UC Berkeley IURD WP 2002 - 4(1).

Repolho, H. M., Antunes, A. P. and Church, R. L. 2013, "Optimal Location Of Railway Stations: The Lisbon-Porto High-Speed Rail Line", *Transportation Science*, Vol.47, n. 3, pp. 330-343.

Reusser, D. E., Loukopoulos, P., Stauffacher, M. and Scholz, R. W. 2008, "Classifying Railway Stations For Sustainable Transitions – Balancing Node And Place Functions", *Journal of Transport Geography*, Vol.16, pp. 191–202.

Rostami , J., Sepehrmanesh, M., Gharahbagh , E. A. and Mojtabai, N. 2013, "Planning Level Tunnel Cost Estimation Based On Statistical Analysis Of Historical Data", *Tunnelling and Underground Space Technology*, Vol.33, pp. 22-33.

RSSB, 2011, "*A Guide to RSSB Research In Station Safety*", Rail Safty and Standared Boards (UK). Available at: http://www.rssb.co.uk/Library/research-development-and-innovation/2011-02-guide-station-safety.pdf Accessed 13 January 2015.

Samanta, S. and Jha, M. 2008, "Identifying Feasible Locations for Rail Transit Stations: Two-Stage Analytical Model", *Journal of the Transportation Research Board*, Vol.2063, n. 2063, pp. 81-88.

Samanta, S. and Jha, M. K. 2011, "Modeling a Rail Transit Alignment Considering Different Objectives", *Transportation Research Part A*, Vol.45, n. 1, pp. 31-45.

Samanta , S. and Jha, M. K. 2012, "Applicability of Genetic and Ant Algorithms in Highway Alignment and Rail Transit Station Location Optimization", *International Journal of Operations Research and Information System*, Vol.3, n. 1, pp. 13-36.

Samanta, S., Jha, M. K. and Oluokun, O. C. 2005. Travel Time Calculation With GIS in Rail Station Location Optimization. *Proceedings of the 25th ESRI International User Conference.* San Diego,CA.

Scheiner, J. 2010, "Interrelations Between Travel Mode Choice And Trip Distance: Trends In Germany 1976–2002", *Journal of Transport Geography*, Vol.18, pp. 75–84.

Schöbel, A. 2005, "Locating Stops Along Bus Or Railway Lines: A Bicriteria Problem", *Annals of Operations Research*, Vol.136, n. 1, pp. 211-227.

Schöbel, A., Hamacher, H., Liebers, A. and Wagner, D., 2002, "The Continuous Stop Location Problem In Public Transportation Networks",Technical Report 81, Wirtschaftsmathematik,University of Kaiserslautern, Germany.

Schöbel, A. and Scholl, S. 2005, "Line Planning with Minimal Traveling Time", In ATMOS 2005: 5th Workshop on Algorithmic Methods and Models for Optimization of Railways. pp. 1-16.

Smith, J. J., Gihring, T. A. and Litman, T., 2015, "*Financing Transit Systems Through Value Capture: An Annotated Bibliography"*, Victoria Transport Policy Institute. www.vtpi.org/smith.pdf.

Špačková, O., Šejnoha, J. and Straub, D. 2013. Tunnel Construction Time And Cost Estimate:From Deterministic To Probabilistic Approaches Prague. *12th International Conference Underground Construction.* Czech Republic.

Srinivas, M. and Patnaik, L. M. 1994, "Adaptive Probabilities of Crossover and Mu tation in Genetic Algorithms", *IEEE Transactions on Systems, Man and Cybernetics*, Vol.24, n. 4, pp. 656-667.

Stringham, M. G. P. 1982, "Travel Behavior Associated With Land Uses Adjacent To Rapid Transit Stations", *ITE journal*, Vol.52, n. 4, pp. 18-22.

Sung, H. and Oh, J.-T. 2011, "Transit-Oriented Development in a High-Density City: Identifying Its Association with Transit Ridership in Seoul, Korea", Cities, Vol.28, pp. 70-82.

Tabassum , M. and Mathew, K. 2014, "A Genetic Algorithm Analysis towards Optimization solutions", *International Journal of Digital Information and Wireless Communications*, Vol.4, n. 1, pp. 124-142.

Tfl, 2014, "*Benchmarking and Financial Planning at TfL*", Transport For London:Finance and Policy Committee, UK. Available at: http://content.tfl.gov.uk/fpc-20141014-part-1-item-09-benchmarking-financial-plan.pdf Accessed 12 July 2015.

The Northern Way, 2009, "*City Relationships: Economic Linkages In Northern City Regions Tyne & Wear City Region*", UK. Available at: http://www.centreforcities.org/publication/city-relationships-economic-linkages-in-northern-city-regions/ Accessed 19 September 2016.

The Regional Ptolemy Model of the East Midlands Region Model Development and Validation Report, 2010, "*LeicesterShire County Council*", UK. Available at: www.leics.gov.uk/ptolemy_mdvr__rev4.pdf Accessed 13 October 2015.

Van Nes, R. and Bovy, P. H. L. 2000, "Importance of Objectives in Urban Transit-Network Design", *Transportation Research Record, 1735*, pp. 25-34.

Vuchic, V. R. 1981, *Urban Public Transportation System and Technology*, Prentice-Hail, Englewood Cliffs, USA.

Vuchic, V. R. 2005, *Urban Transit: Operations, Planning & Economics*, John Wiley & Sons , New Jersey, USA.

Vuchic, V. R. 2007, *Urban Transit System and Technology*, John Wiley and Sons, New Jersey, USA.

Vuchic, V. R. and Newell, G. F. 1968, "Rapid Transit Interstation Spacings for Minimum Travel Time", *Transportation Scince*, Vol.2, pp. 303-309.

Vuchic, V. R. 1969, "Rapid Transit Interstation Spacings for Maximum Number of Passengers", *Transportation Science*, Vol.3, n. 3, pp. 214-232.

Webtag Data Book, 2015, "*Department for Transport*", UK. Available at: https://www.gov.uk/government/publications/webtag-tag-data-book-december-2015 Accessed 9 January 2016.

Webtag: Tag Unit M2 Variable Demand Modelling, 2014, "*Department for Transport*". Accessed 13 January 2014.

Wirasinghe, S. C. and Vandebona., U. 1987. Some Aspects of the Location of Subway Stations and Routes. *Presented at the Fourth International Symposium on Locational Decisions (ISOLDE).* Namur, Belgium.

Yoh, A., Haas, P. and Taylor, B. 2003, "Understanding Transit Ridership Growth: Case Studies of Successful Transit Systems In The 1990s.", *Transportation Research Record*, n. 1835, pp. 111–120.

Zemp, S., Stauffacher, M., Lang, D. J. and Scholz, R. W. 2011, "Classifying Railway Stations For Strategic Transport And Land Use Planning: Context Matters!", *Journal of Transport Geography*, Vol.19, pp. 670-679.