ABSTRACT

Title: OPTIMIZATION OF STATION LOCATIONS AND
                                TRACK ALIGNMENTS FOR RAIL TRANSIT LINES

                                Xiaorong Lai, Doctor of Philosophy, 2012

Directed by: Dr. Paul Schonfeld, Professor
                                Department of Civil & Environmental Engineering

Designing urban rail transit systems is a complex problem, which involves the determination of station locations, track geometry, right-of-way type, and various other system characteristics. The existing studies overlook the complex interactions between railway alignments and station locations in a practical design process. This study proposes a comprehensive methodology that helps transit planners to concurrently optimize station locations and track alignments for an urban rail transit line. The modeling framework resolves the essential trade-off between an economically efficient system with low initial and operation cost and an effective system that provides convenient service for the public. The proposed method accounts for various geometric requirements and real-world design constraints for track alignment and stations plans. This method integrates a genetic algorithm (GA) for optimization with comprehensive evaluation of various important measures of effectiveness based on processing Geographical Information System (GIS) data.

The base model designs the track alignment through a sequence of preset stations. Detailed assumptions and formulations are presented for geometric requirements, design constraints, and evaluation criteria. Three extensions of the base model are proposed. The first extension explicitly incorporates vehicle dynamics in the design of track alignments, with the objective of better balancing the initial construction cost with the operation and user costs recurring throughout the system's life cycle. In the second extension, an integrated optimization model of rail transit station locations and track alignment is formulated for situations in which the locations of major stations are not preset. The concurrent optimization model searches through additional decision variables for station locations and station types, estimate rail transit demand, and incorporates demand and station cost in the evaluation framework. The third extension considers the existing road network when selecting sections of the alignment. Special algorithms are developed to allow the optimized alignment to take advantage of links in an existing network for construction cost reduction, and to account for disturbances of roadway traffic at highway/rail crossings. Numerical results show that these extensions have significantly enhanced the applicability of the proposed optimization methodology in concurrently selecting rail transit station locations and generating track alignment.

OPTIMIZATION OF STATION LOCATIONS AND TRACK ALIGNMENTS FOR
RAIL TRANSIT LINES

By

Xiaorong Lai

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of philosophy
2012

Advisory Committee:
Professor Paul Schonfeld, Chair/Advisor
Professor Ali Haghani
Professor Subramanian Raghavan
Professor Manoj K. Jha
Assistant Professor Cinzia Cirillo

Dedication

To my wife, Ying Liu

And to our wonderful daughter, Alyson Yuxuan

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my advisor, Professor Paul Schonfeld, for his academic guidance, encouragement and support over the past five years. I am very grateful for his willingness to meet even after the normal working hours, and I have greatly benefited and inspired from discussion with him. I feel very fortunate to have been his student.

I also wish to thank my committee members, Professor Ali Haghani, Professor Subramanian Raghavan, Professor Manoj Jha and Professor Cinzia Cirillo, for their precious comments and suggestions on my dissertation. Special thanks go to Professor Manoj Jha in providing valuable thoughts and references throughout my research.

I would also like to thank my parents and three elder sisters for all the support they have provided me over the years.

Finally, and most importantly, I would like to thank my wife, Ying Liu, for always standing by my side through the good times and the bad times. It was her love, encouragement, and support that helped me the most,

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Research Background

Urban rail transit is an all-encompassing term for various types of local passenger rail systems serving urban and/or suburban areas. These systems may differ in many aspects, such as track technology, train capacity and right-of-way type, but they share the similarity of operating trains along tracks of a fixed alignment and, compared with intercity rail, serving shorter corridors with more densely spaced stations. Light rail transit and rail rapid transit (subway or metro) are the two most commonly used types of urban rail transit systems.

During recent decades, urban development, travel demand increases and the growing need for high-performance transit that is separate from frequently congested urban streets have stimulated the construction or expansion of urban rail transit systems in many US cities [Vuchic, 2005]. Such projects generally require substantial investments and exert a permanent impact on travel patterns and even urban land-use patterns. Thus, their actual implementation requires sophisticated planning and evaluation efforts, as mistakes in the design will result in inefficiencies that are difficult to correct [Vuchic, 2005].

Designing urban rail transit systems is a very complex problem, which involves the determination of station location, track geometry, right-of-way type, and various other system components. The design represents the essential trade-off between an economically efficient system with low construction and operation cost

and an effective system that can provide convenient service for the public. The design process is also constrained by many factors, including project budget, design specifications, and various local conditions such as travel demand patterns, topography, existing street networks as well as safety, environmental, and public policy issues.

The current design practice of urban rail transit systems is mostly approached empirically, depending heavily on the planners' judgment and with little theoretical basis. With a trial-and-error process, planners first develop one or more candidate plans that can satisfy the design constraints and meet the performance requirements. These alternatives are then extensively evaluated with respect to the preset criteria or measurements of effectiveness. Such a design process is not only time consuming due to its repetitive manual processes, but also cannot guarantee that its result is even close to the optimal alternative(s).

The theoretical studies in planning urban rail transit networks and facilities have largely ignored the need for detailed network design. Among the rather limited number of publications on optimizing urban rail transit designs, many are difficult to apply in real world applications due to the neglect of practical design constraints, the adoption of unrealistic assumptions, or the requirement of significant computational efforts. Greater attention is needed to bridge the present gap between actual design practice for specific areas and applied research for development, analysis and evaluation of transit system designs.

This dissertation focuses on the design of two essential system components of urban rail transit lines, i.e., the location of rail transit stations and the alignment of

railway tracks between stations. Both components are indispensable to fulfill the basic functions of rail transit operations, namely, the collection/distribution of passengers and their transport over a distance. Both components involve extensive infrastructure and thus are much more expensive to change after construction, or in other words, more "permanent", compared with other aspects of urban rail transit systems (e.g., operation frequency or train capacity).

The proposed methodology incorporates both station locations and railway track alignments into the optimization procedures, by accommodating multiple system objectives, formulating various design constraints, integrating the analysis models with a GIS database and developing effective solution search methods. It is expected that transit planners may greatly benefit from the proposed methodology, with which they can conveniently generate well-optimized candidate alternatives in an efficient way.

1.2 Research Objectives

The overall objective of this dissertation is to develop an effective methodology that helps transit planners to produce optimized alternative designs of station locations and track alignments for an urban rail transit line. To achieve this objective, the proposed methodology should have the following features:

- Evaluate and optimize rail transit routes and stations based on the multiple effectiveness measures, objectives and constraints which are pertinent

- Readily fit in practical design circumstances and incorporate potential predefined design components

– Realistically represent the track alignment and station layout

– Effectively account for various local conditions

– Properly evaluate different design alternatives

– Efficiently generate good design alternatives

– Conveniently access given information and demonstrate resulting designs.

In response to the aforementioned modeling features, this study pursues several research goals listed below:

1. Development of a modeling framework that satisfies the need of transit planners.

2. Formulation of appropriate performance measures for evaluating the design alternatives, which account for various costs associated with system construction/ operations, and the potential cost savings by introducing the rail transit line.

3. Generation of track alignment and stations plans that meet various geometric requirements.

4. Incorporation of real-world design constraints regarding geographical restrictions and other local conditions.

5. Development of effective solution search methods to enhance computational efficiency and solution quality of the optimization process.

6. Conduct of case studies to examine the effectiveness of the proposed methodology.

7. Development of a computer program to integrate the proposed optimization methodology with a Geographical Information System and with user-friendly graphical interfaces.

This study features a modeling framework for resolving the essential trade-off between an economically efficient system with low initial and operation cost and an effective system that provides convenient service for the public. The proposed method accounts for various geometry requirements and real-world design constraints for track alignment and stations plans. This method integrates a genetic algorithm (GA) for optimization with comprehensive evaluation of various important effectiveness measures based on processing Geographical Information System (GIS) data.

1.3 Thesis Organization

The organization of this dissertation is as follows:

Chapter 2 presents a comprehensive review of existing studies in locating rail transit stations and designing railway track alignment. The review focuses on four aspects of these studies: the representation of stations and/or alignment with geometric objects, the selection of criteria to evaluate design alternatives, the incorporation of geometry requirements and other design constraints, and the use of methodologies or procedures to generate design alternatives. The review aims to find essential design issues in planning an urban rail transit system, and identify those aspects that need further improvement in the design process.

Chapter 3 presents the base model and solution algorithm for integrated optimization of rail transit station locations and track alignment. The model aims to design track alignments between a sequence of stations that planners preset at major demand points and/or transfer centers. The chapter details the assumptions and formulations of the optimization model, paying special attentions to evaluation criteria, geometry requirements and design constraints which are unique. This chapter also illustrates a solution procedure based on a Genetic Algorithm that is designed to account for the non-differentiable cost functions and complex geometry requirements of a real-world rail transit alignment.

Chapter 4 illustrates the development of a computer program and an extensive case study, using the Baltimore Red Line as an example. The program embeds the proposed optimization algorithm and a Geographical Information System. The case study aims to demonstrate the efficiency of the proposed optimization algorithm and provide some insights of the problem with extensive sensitivity analysis. Based on the sensitivity results, this chapter also presents an improvement of the proposed algorithm by incorporating additional decision variables.

Chapter 5 presents the first extension of the base model, which explicitly incorporates vehicle dynamics in the design of track alignments. The proposed model aims to reliably estimate travel time and energy consumption, and to achieve the desirable trade-off between the initial cost incurred at the onset of the project and the operation and user costs recurring throughout the system's life cycle.

Chapter 6 presents the second extension of the base model, which concurrently optimizes rail transit station locations and track alignment. The chapter

details the methodology for generating the candidate pool of potential rail transit stations, the embedded rail transit demand forecasting module, and its interaction with the concurrent optimization model. The case study aims to demonstrate the advantage of concurrent optimization over the two-stage optimization.

Chapter 7 presents the third extension of the base model, which takes the existing network into consideration. Special techniques and algorithms are developed to allow the optimized alignment to use existing network links for construction cost reduction, and to account for disturbances of roadway traffic at highway/rail crossings. The proposed model has the ability to search for the best combination of existing network links and new alignment segments in order to minimize the total cost.

Chapter 8 summarizes the tasks completed in this dissertation and potential topics for future research.

Chapter 2: Literature Review

The literature review for this study includes three sections, corresponding to three aspects of research associated with the design of urban rail transit systems. The first section examines the theoretical work in determining rail transit track alignments. It also covers various highway alignment optimization models, as the two alignment problems share similarities in many aspects. The review pays particular attention to various design constraints and the way they are incorporated in the formulation. Section 2 examines various analytical models for selecting the locations of transit stations, with special focus on their assumptions, objective functions, and design constraints. As reviews in the first two sections demonstrate the need for criteria with which to evaluate track alignments or select station locations, Section 3 presents a brief review on the variety of costs and potential benefits associated with the construction and operation of urban rail transit systems. A summary of findings from the literature review is provided at the end of this chapter.

2.1 Design of Track Alignment

The route upon which a train travels and the track is constructed is defined as an alignment [AREMA, 2004]. An alignment consists of two components: the horizontal alignment defines physically where the track goes (the XY plane); and the vertical alignment defines the elevation along the track (the Z component). Research on the classic track alignment problem, which tries to establish track alignment

between two given points, dates back 150 years [Vuchic, 2005]. The challenge has been, and still is, the efficient selection of the optimal route alternative(s) subject to a complex set of constraints, including train-track dynamics, operation safety, construction and maintenance costs, short and long-term traffic impacts, environment restrictions, and other political or economic concerns.

This section reviews existing alignment optimization methodologies and ends with a summary of general design considerations that should be accounted for in the track alignment design of urban rail transit routes. It should be noted that track alignment optimization models are really rare for rail transit system. Studies on track alignment have focused mainly on the evaluation of specific alignment components, not on the automatic generation of track alignments. Meanwhile, studies on optimization of transit routes are mostly based on simple assumptions of track alignment. For example, Chien and Schonfeld [1997] assumed a grid transit system where transit routes were either horizontal lines or vertical lines and the construction costs solely depended on the zone the route traverses. Lai et al. [2002] assumed that the light rail route would follow the existing roadway network and no construction costs were taken into account.

Hence, the review in this section will not be limited to models for rail track alignment but extended to those for highway alignment, as the two design problems share many similarities: they both search for a consecutive series of spatial elements satisfying certain geometry requirements; they both involve substantial investment; they are both restricted by topological, land use and environmental features; and they both deal with large amounts of spatial data and incur complex computational

problems. Review of these highway alignment optimization models is expected to provide some insights on the selection of optimization objectives, the representation of the route alignment, the consideration of design constraints and the computation of optimal solutions for the track alignment optimization problem.

2.1.1 Alignment Optimization Models

Jong [1998] and Jha [2000] classified earlier highway alignment optimization models into several general groups, mainly based on problem formulation and solution algorithm:

- Enumeration. Similar to the engineering practice, this method compared all the possible alignment alternatives to find the optimal one. One example is Easa's work [Easa, 1988] on the vertical alignment design. The major limitation of this method is its inefficiency, especially as there are usually a large numbers of feasible alternatives in practice.

- Calculus of variations. Treating the alignment as spatial curves following a predefined surface, this method tried to minimize the cost that was represented as integrals of the curve function. Examples include Howard, et al. [1968] and Shaw and Howard [1981, 1982] for horizontal alignment optimization, as well as for vertical alignment optimization. This method can provide a continuous and global optimal alignment. The major concern is how to represent complex geometry requirements, and how realistic is its assumption of a continuous cost function, which is hard and

sometimes impossible to define in practice when location-specific costs (e.g. right-of-way costs) are involved.

- Numerical search. Also using a continuous search space, this method applied numerical search technique to find the optimal alignment with the minimal cost. Examples include Hayman [1970] and Goh et al. [1988] for vertical alignment design and Chew et al. [1989] for concurrent optimization of horizontal and vertical alignment. This method allows more flexibility by using a continuous search space, but it cannot guarantee a global optimal solution. Besides, a numerical search approach generally requires a differentiable objective function and thus cannot handle discontinuous cost items such as location-based right-of-way cost.

- Linear programming. Assuming the alignment follows pre-specified function form(s), this method applied linear programming techniques to optimize the coefficients of the function so as to minimize the cost. Examples include the work of Moreb [1996] and ReVelle et al. [1997] in vertical alignment design. This method provides formulations that are easy to solve; however, it is usually hard to justify the selected function forms.

- Network optimization. This method represented the alignment as a consecutive series of predefined arcs and tried to minimize the total arc cost. Examples include Turner [1971, 1978], Athanassoulis and Calogero [1973], Parker [1977], and Trietsch [1987a, b] for horizontal alignment optimization, as well as for vertical alignment optimization. This method

can provide a global optimal alignment with well-developed solution techniques. The major concern is how to pre-determine those candidate arcs and their associate costs in engineering practices, and how to enforce various geometry constraints.

- Dynamic programming. This method divided the alignment into segments (stages), and from the start point, each segment could end at several candidate points (state). Examples include Trietsch [1987a] for horizontal alignment design, Puy Huarte [1973], Murchland [1973], Goh et al. [1988] and Fwa [1989] for vertical alignment design, as well as Hogan [1973] and Nicholson [1976] for concurrent optimization of horizontal and vertical alignment. This method can also provide a global optimal alignment with readily available solution techniques, but the alignment is mostly piecewise linear in nature. Other concerns are how to enforce the geometry constraints and how to define the stages and states in practice to achieve a proper trade-off between alignment accuracy and computational burden.

The remaining section reviews in more detail those more recent studies regarding railway and highway alignment optimization.

HAO Models

A research team at the University of Maryland has proposed a series of GA-based highway alignment optimization models. Jong [1998] and Jong and Schonfeld [2003] first demonstrated the concept of applying a GA to build the horizontal and

vertical alignment concurrently. The objective function included three types of construction costs (i.e., location dependent cost, length-dependent cost and earthwork cost), three types of user costs (i.e., vehicle-operating costs, travel time cost and accident costs) and penalties for design constraint violations. The proposed approach first generated the location of PIs and VPIs on a fixed number of cutting plans. Two iterative procedures were then used to fit the minimum radius circular curve at each PI for horizontal alignment and minimum length parabolic curve at each VPI for vertical alignment.

Jha [2000], Jha and Schonfeld [2004] and Jha et al. [2006] extended the HAO models by integrating a GIS to better accommodate the complex topological and environmental features. Kim [2001] and Kim et al. [2004] developed methods for incorporating the cost of major structures (i.e., bridges and tunnels) in the model objective function. Kang et al. [2007] further improved the GA based solution algorithm by introducing the Feasible Gates approach. This approach, reflecting the natural restrictions of certain topological features (e.g., flood plains and wetlands) on alignment design, can greatly reduce the solution space and increase solution efficiency. Kang [2008] incorporated the traffic assignment process into the HAO model and discussed the optimization problem of adding a new highway to an existing road network. Jha et al. [2007] extended the HAO models to railway alignment optimization. That model adopted different design criteria and cost functions, but the methodology stayed mostly the same.

Model of Fwa et al [2002]

Fwa et al. [2002] proposed a model for solving the vertical alignment of a highway segment given the horizontal alignment. The objective function included both earthwork cost and pavement cost. The vertical alignment was represented with a series of VPI points on vertical grid lines. The model introduced three design constraints: critical length of grade control, fixed-elevation points, and nonoverlap of horizontal and vertical curves. However, these constraints were realized simply by adding a fixed large penalty for violation in the proposed GA based solution algorithm. Such a constant penalty function may lead to large unsmooth steps during the optimization process, and thus fail to yield optimal solutions [Kang, 2007].

Model of Cheng and Lee [2006]

Cheng and Lee [2006] proposed an approach for solving the three-dimensional alignment of a highway segment. The objective included minimizing the weighted penalties for violating control point/restricted area requirements, minimizing weighted highway length as horizontal alignment cost, and minimizing the overall costs associated with vertical alignment. The target horizontal alignment consisted of tangents, circular curves, and clothoid-type transition curves, whereas the target vertical alignment was represented by grades and vertical curves. To satisfy various geometric requirements such as minimal horizontal curve radius/length, maximal/minimal vertical curve length and maximal slope, the proposed approach optimized the alignment with an iterative process. Within each iteration, a new horizontal alignment was generated from previous alignment by slightly adjusting the

points of intersection (PIs) and then inserting curve elements between neighboring PIs. A corresponding vertical alignment was then obtained by solving a series of linear mixed integer models [Lee and Cheng, 2001a, 2001b].

The key contributions of this model are its use of transition curves to realistically represent the curved sections of horizontal alignments and its consideration of speed reduction constraints for heavy vehicles in designing vertical alignments. Its limitations include the neglect of right-of-way cost in the objective function and the simplified assumption of vertical construction cost as a linear function of the VPI elevation.

Model of de Smith [2006]

De Smith [2006] proposed a general approach for finding the optimal alignment for roads, railroads, and pipelines. The optimal alignment was generated via three steps: 1) find the shortest alignment that satisfies gradient constraints in a lattice approximating the existing ground; 2) smooth the horizontal alignment to meet curvature constraints, and 3) smooth the vertical alignment to meet gradient constraints. This approach can incorporate infeasible areas and location-based costs; however, all boundaries are assumed to be straight lines parallel to the line between the start and endpoints of the alignment. Besides, this approach cannot guarantee the resulting alignment is optimal.

2.1.2 Design Considerations in Determining Track Alignment

Although addressing different alignment components and employing various assumptions, the aforementioned alignment optimization models have identified some common issues that must be dealt with in designing track alignment.

Decision Variables

In the design practice of track alignment, the guidelines developed over the past two centuries generally cover three alignment components [AMTRAK, 2003; AREMA, 2004]:

- Horizontal alignment is the projection of the three-dimensional rail track onto the two-dimensional XY plane. It consists of a series of straight sections of track, referred to as tangents, connected by simple, compound, reverse, and/or transition (spiral) curves.

- Vertical alignment defines the elevation of every point along the horizontal alignment. It consists of a series of straight lines, called grades, which join to each other by vertical curves (almost always parabolic in nature).

- Superelevation is the rise of the outside rail in a curve by rotating the track structure about the point of rotation (typically the inside rail). It is provided to counteract or partially counteract the centrifugal force due to curvature and speed.

The design of track alignment must account for all these components. Yet existing alignment optimization models all exclude superelevation from their considerations, although it may impact riding comfort and safety. This can be

explained in two ways. First, superelevation does not significantly impact the construction cost, especially for railway tracks with limited cross section width. Also, superelevation can be indirectly reflected by other geometric components, such as curve radius and tangent length

Evaluation Criteria

Vuchic [2005] summarized three major categories of objectives in rail transit systems planning.

－ Perform maximum network passenger attraction,

－ Achieve maximum operating efficiency, and

－ Create positive impacts.

Existing alignment optimization models generally focus on the minimization of system costs, which is only one aspect of the system efficiency. Passenger attraction is mostly left out of the picture, assuming it will be captured either earlier in the selection of the two end terminals or later in locating stations along the alignment.

Geometry Constraints

Geometry constraints weigh more heavily for rail track design than for highway design for three reasons [AREMA, 2004]. First, trains are operated with automatic guidance mechanism along the fixed track alignment. The horizontal movements of trains are beyond operators' control but rely on the alignment only. Secondly, the ratio of locomotive power to vehicle mass is significantly less than for

automobiles, which leads to lower acceleration/deceleration and thus much longer response time/distance. Finally, trains have extremely long and thin dimensions, which may cause various internal forces undesirable for operation safety. These concerns result in much stricter requirements on geometric design for track alignment.

For horizontal track alignment, the most important design constraints include maximum curvature of simple curves, minimal length of tangent tracks between adjacent reverse curves, and minimum length of each element in a compound curve [Hay, 1982; Black, 1995; AMTRAK, 2003; AREMA, 2004]. They were proposed mainly for smoother and safer operation of railway vehicles, whereas for vertical alignment, the most important design constraints were proposed mainly for better locomotive/brake performance and safer train operation. The constraints included maximal grades and minimal lengths of vertical curves [Hay, 1982; Black, 1995; AMTRAK, 2003; AREMA, 2004].

Also notable is that a few studies have emerged in recent years on the geometry requirements of integrated horizontal and vertical alignment. For example, Smith and Lamm [1994] addressed the 3D nature of the highway alignment in designing aesthetically pleasant highways. Sanchez [1994] studied the sight distance on interchange connectors in 3D combined projections. Hassen et al. [1997] also studied the effect of considering 3D alignment on design requirement for sight distance using a finite-element-technique based analytical model to compute sight distance. Kuhn and Jha [2011] proposed a methodology to check the safety-related and esthetic shortcomings of a 3D alignment when its horizontal projection, vertical

projection and cross-sections are processed separately and then superimposed. Results from the latter three studies showed that the 3D design requirements may differ significantly from those in separate 2D projections. Jha et al. [2011] proposed a 3-D design methodology that was based on the development of the road surface, the virtual field-of-view surface and a virtual line-of-sight plane. An algorithm was also proposed to calculate sight distance from the three developed surfaces. However, there are no actual 3D-based design standards available in the literature.

Environmental and Topography Constraints

Another type of constraints considered in alignment design is the environmental and topography constraints. Unlike the geometry constraints that focus on the track-vehicle system itself, this type of constraints tries to account for external factors, such as:

- Topography features that restrict the possibilities of alignment design, such as hilly terrain, valleys, rivers, or lakes

- Environmentally sensitive areas that the alignment should bypass, such as wetlands and historic districts

- Fixed points (or areas) through which the alignment must pass

- Existing roadway network that may provide right-of-way and thus reduce construction cost. It may also create difficulty in crossing and access design, and cause unpleasant riding environment of intensive noise and air pollution.

Vehicle Dynamics

Another trend in geometry design of track alignments has been the incorporation of vehicle dynamics. Kim and Schonfeld [1997] examined the benefits of dipped vertical alignment for rail transit, where the vertical profile starts getting lower upon leaving a station and then gradually picks up elevation before the next station. The authors set up a simulation model using basic equations of dynamics, and demonstrated the benefits of such vertical alignment in reducing both propulsive and braking energy. Klauder et al. [2002] simulated the train-track dynamics of a rail vehicle operating over two railroad curve transition spiral shapes and compared their dynamic performance. Kim and Chien [2010] developed a time-driven train performance simulation (TPS) model to emulate the movement of a train, calculate energy consumption, and estimate travel time, considering various vertical track alignments and operational controls. Kufver [1997, 1998] and Kufver and Andersson [1998] in a series of studies in the late 1990s considered vehicle reactions in alignment optimization, but the work was more focused on ride comfort and single alignment components such as circular or transition curves. Using a deterministic simulation model based on basic kinematics and resistance relations, Yeh [2003] proposed a model to jointly optimize vertical alignment and operating characteristics such as speeds and coasting distances. However, the model only considered simple dipped profiles and one-directional operation between two stations. The model also assumed the construction cost was not affected by vertical profiles.

2.2 Selection of Rail Transit Stations

Rail transit stations are points along rail transit lines where trains stop for passengers to board and/or alight. Unlike bus stops that can be easily relocated, rail transit stations are permanent structures that involve major investment and often have strong impacts on their surroundings [Vuchic, 2005]. Locations of rail transit stations also significantly impact passenger attraction as well as operations of rail transit system, such as operating speed of trains, travel time, riding comfort and operating costs. These facts indicate that determining the locations of rail transit stations is a critical part in designing a rail transit system.

The remaining sections first give a detailed review of the modeling efforts in optimizing the locations of rail transit stations. The review ends with a discussion of various design considerations in selecting rail transit station locations. As the review shows that potential passenger attraction plays an important role in the station selection process, a subsection is followed to brief the research work on ridership forecasting of rail transit systems.

2.2.1. Transit Station Location Models

The existing models for optimizing rail transit stations fall into two general categories. The first category locates rail transit stations along a given rail transit alignment, whereas the second category, without knowing the alignment, tries to select stations from candidate sites and decide the sequence of selected stations. Models in the second category are sometimes referred to as integrated optimization

models of station and alignment, although they cannot yield the real track alignment that satisfies various geometric constraints.

Category I Models

Early studies in this category tried to find interstation spacings along a given rail transit line where people commute to a single point. Vuchic and Newell [1968] reviewed the work of their predecessors and criticized their assumption of uniform population density. In response, their analysis took into account passenger distribution along the line as well as dynamic characteristic of the train and intermodal transfer time at stations. With respect to the objective of minimum passenger travel time, the spacings were functions of the ratio between the number of passengers traveling on the train and those wanting to board or alight. Similar to the earlier studies, this work did not consider access time and competitive transportation modes. The authors later also noted that the most desirable spacing were often greater if accounting for practical considerations, such as maintaining high operating speed, attracting more passengers traveling longer distances and achieving a lower cost [Vuchic, 2005]. In an extended effort, Vuchic [1969] studied the station spacing to achieve maximum ridership. With the same basic model, additional assumptions were employed such as uniform population distribution along the line and use of the same alignment for the competitive system. Another extension was provided by Kikuchi and Vuchic [1982], who developed a theoretical model to calculate the optimal station spacing and vehicle stopping policy for a rail transit line. The objectives were minimum user travel time and minimum total cost.

Considering non-uniform many-to-many travel demand, Wirasinghe and Ghoneim [1981] built an analytical model to determine the optimal spacing of bus stops along a local bus route. The study used partial differential equations to minimize the passenger travel-time cost plus operating costs, assuming that user access time toward the bus route is independent of station locations and that a cumulative boarding/alighting function along the bus route is available exogenously.

Wang et al. [2004] tried to locate and price one single park-and-ride facility in a linear monocentric city, which is in fact equivalent to the transit route in early studies. This work yields analytical expressions of the optimal PNR location and parking charges for maximizing profit and minimizing social cost respectively. However, the results have only limited practical applications due to the use of simplified assumptions, such as uniformly distributed residences from the center to the exogenous city boundary, a congestion-prone highway and a congestion-free railway accessible at all points, all trips from home to city center, and deterministic mode choice based on user equilibrium.

Laporte et al. [2002] sought to locate stations on a fixed rail transit alignment so as to maximize the ridership, subjected to inter-station spacing constraints. Assuming the percentage of captured travelers decreases with their access distance, the study estimated the ridership of each potential station by triangulations of census tracts and approximation of access distance as predefined weighted norm. To deal with the adjoining catchment area between neighboring stations, it simply assumed that passengers always choose the closest station. The station locations can then be obtained as the longest path on a directed graph, which only contains links between a

pair of candidate stations if they meet the inter-spacing restrictions.

Similarly using GIS tools, Samanta et al. [2005] used an ant algorithm to optimize station locations along a rail transit line so as to minimize the overall system cost (i.e, capital cost, operator cost, and user cost). Travel times to proposed rail stations were calculated using actual road network, but only from centroids of residential locations.

Category II Models

The second category of station location models generally start by tentatively designating a large number of potential station locations and then search for a consecutive series of links between these stations as the final alignment.

Early models in this category included two bicriterion mathematical programming models developed by Current et al. [1985, 1987]. The Maximum Coverage Shortest Path model tried to minimize total construction cost and to maximize total demand covered, while the Median Shortest Path model involved the minimization of total construction cost and the maximization of path accessibility (i.e. the total weighted distance that nodal demand must travel to reach the closest station). Both models were based on the simple assumptions of given link cost and fixed radius for station coverage.

Dufourd et al. [1996] addressed the problem of locating a fixed number of stations for a rapid transit line with known terminus on a grid network. The model was formulated as a longest-path problem to maximize the total population covered by stations, subject to interstation spacing constraints. The calculation of station

coverage was based on simplified assumptions that 1) stations do not have overlapped catchment areas; and 2) each discretized demand in a station's catchment area is assigned with non-increasing weight based on its distance to the station on a Manhattan metric. To account for the side constraints, the authors developed a Tabu search heuristic that basically proceeds with neighborhood search and allows intermediate deteriorating solutions so as to prevent local optimum. Similarly with a population coverage objective and station inter-spacing constraints, Bruno et al. [2002] proposed another heuristic to gradually extend a partial alignment by locating one location at a time while ensuring the interspacing constraints.

Samanta and Jha [2008] proposed a two-stage analytical model for locating rail transit stations. The upper model, embedded within a GIS, identified feasible station sites to avoid interference with existing road network and built-up areas (e.g., residential neighborhoods and business establishments). The lower model applied a GA algorithm to seek the best set of stations for minimization of the overall system cost, subject to interstation spacing constraints. Among the overall system cost, the capital cost of stations only considered the right-of-way cost. The operator cost was assumed to be length-dependent and vary linearly with the distance between neighboring stations. Samanta [2008] and Samanta and Jha [2011] furthered their research to use different objective functions of demand and cost. Following the method proposed by Lee and Vuchic [2005] to consider variable demand in transit network design, the model also ran an iterative modal split process to reflect the impact of station locations/sequence on rail transit demand of a many-to-one travel pattern. Focused more on the solution algorithms, Samanta and Jha [2012] compared

25

Genetic Algorithm and Ant Algorithm in solving highway alignment models and rail transit station location optimization models.

None of the aforementioned models provided a satisfactory description of real-world situations. They either overlooked the travelers' mode and route choice behavior when alternative transportation systems other than rail transit are available, or simply assumed those behaviors are captured with the input data available exogenously. Some approaches were thus developed to address these issues.

Assuming predefined origin-destination demand, Bruno, et al. [1998] developed a bicriterion approach to locate stations between two given terminals. With the objective of minimizing the overall construction cost and the weighted travel cost incurred, the approach started by identifying K shortest paths in construction costs on the transit network. For each path, a bi-modal network was built to represent the private and hybrid pedestrian-transit alternatives. Assuming each user chooses the least-cost route, the approach easily calculated the travel cost incurred and then selected the efficient solutions using the dominance relationship. The model required extensive data input, such as a known OD demand, fixed travel cost on private network, and given construction cost and travel cost for transit links between candidate transit stations. Besides, no station construction cost and spacing constraints were considered.

Without assuming a predefined OD demand, Laporte, et al. [2005] described two greedy heuristics to choose stations among a set of candidate locations for a rapid transit line. The objective was to maximize the total OD demand covered by the alignment, which was estimated with a modified station catchment model. The model

first derived a trip coverage matrix for any two stations using gravity models, where each element in the matrix represents the demand they caught between each OD pair. The model then calculated OD demand via transit with a simple logit model, assuming fixed cost for complementary traffic mode. This study only considered the maximum station spacings constraints. More critically, the proposed approach for estimating OD demand between stations had neither a theoretical base nor practical verification/validation.

Marin et al.[2009] addressed the station location problem for a rail transit network. The optimization objective was to maximize the transit demand and minimize the total travel time in the complementary network, subjected to budget, user's behavior, and network design constraints. Between each OD pair, the proposed model formulated the complementary mode as a fictitious link with fixed cost, and defined mode splitting with a predefined Logit model. As the two objectives both favored higher transit demand and thus a more efficient transit network, the route choice was simply modeled by the flow conservation law in transit network. The model also introduced a new constraint to limit the number of routing intersections in transit network so as to restrict the number of transit lines. With many inputs required (e.g., rail line/station construction cost and link travel time) and an approximation of the Logit function as piecewise linear, the model can be transformed to a linear integer programming model and solved with the commercial software CPLEX.

There are some other studies about locating rail transit stations/park-and-ride facilities. Horner and Grubesic [2001] developed a GIS-based method to generate a suitability index for potential sites of park-and-ride facilities along urban rail lines.

The gravity-based index, representing demand within each site's catchment area, was calculated with derived rail demand and network travel times, taking into account the competition among alternative modes and candidate sites. Farhan and Murry [2003] developed a GIS-based approach to delineate catchment areas for park-and-ride facilities, which simultaneously accounted for park and ride facility accessibility and user travel direction. Faghri et al [2002] developed a Knowledge-Based Geographic Information System to evaluate candidate locations of Park-and-Ride facilities, based on predefined criteria and weight of each criterion. The system excluded travel demand characteristics in the evaluation and could only serve the urban areas with congestions caused by heavy inbound commuting traffic. Similar work was found in Wey and Chang [2007], which tried to select Joint Development Stations for a Mass Rapid Transit (MRT) system in Taiwan. Their approach, also based on predefined criteria, applied some analytic process to generate weights from experts' opinions and to make the selection. These studies are more focused on the evaluation and comparison of individual sites, rather than using optimization models to select a set of rail transit stations.

Design Considerations in Locating Rail Transit Stations

Although addressing different problems and employing various assumptions, those previous models have identified some common issues that must be considered in selecting rail transit station locations.

Candidate Station Sites - Though some work assumed the candidate sites for rail transit stations could be anywhere along the rail transit line [Vuchic, 1968,1969]

or in the study area [Dufourd, et al., 1996], most studies agreed that the candidate station sties should be limited to locations satisfying some general requirements, such as topology, existing road network and land availability. Also, to increase the potential ridership of rail transit system, the station locations are usually fixed by the locations of other transportation terminals, major activity centers, college campuses, highway intersections, etc. Thus, further examination is warranted on how we should efficiently generate these candidate sites and effectively embed them in the station location and alignment optimization process.

Selection Criteria - Ridership attraction is the most popular design criterion in locating rail transit stations. It is represented either by the actual number of passengers attracted to rail transit mode or by weighted area coverage (i.e. potential passenger coverage). Some other measurements were also applied in the literature, such as the cost of the rail transit system (i.e., investment cost, operating cost, and/ or user cost) or even subjective judgments. The individual objectives are not always mutually compatible and thus the search for a compromise based on quantitative measures becomes an issue in the design process.

Station Type - Previous studies all assumed the type of stations they tried to locate were predefined: either pedestrian-orientated or with park-and-ride facilities. Most early works, focused on densely populated urban area or suburban centers, assumed that a significant portion of transit riders access transit services on foot so as to avoid the costs associated with owning/driving/parking a vehicle. Some later studies on locating park-and-ride facilities, on the other hand, paid more attention to suburban areas where many transit riders would access to stations by auto. The two

types of stations target different groups of users, have different attraction radius, involve different land needs, and involve different cost and environmental concerns. Thus the selection of station types is an integral part of the station location problem and deserves further attention.

Operational Properties - Some early models have tried to capture the dynamic characteristics of trains in their analysis, whereas other studies mostly tried to circumvent this issue with some kind of strategies. Examples include the use of the minimal inter-station spacing constraints to restrain the average operation speed [Larpote, et al. 2000], or even the direct use of assumed average travel speed or inter-station travel time. Other operational issues not covered in station location models include fleet size, vehicle scheduling, fare system and etc. Further study is needed on how to effectively integrate such operational properties in the design of rail transit systems.

Geometry Constraints - Existing studies on locating rail transit stations, whether along the rail transit line or not, assumed the station was a single point and enforces no additional geometry constraints. Yet an actual rail transit station has its own layout and elements, for example, the platform alongside rail tracks from which passengers board or alight from trains. For safety and cost considerations, most station design practices require the platform to be located along a minimum length of tangent track whose grade and cross slope do not exceed given thresholds [Davies, 2007; FWTA, 2007; I-70 Coalition, 2008]. It makes more sense to embed such constraints directly in the station location/ alignment optimization process, rather than shift the stations later.

Competitive Modes - To justify the high cost associated with the built and operation of rail transit system, sufficient trips need to be attracted from alternative modes so as to generate revenue and reduce congestion. Existing studies either assumed the mode and route choice behavior were captured with the input data available exogenously, or applied simple models with various assumptions to represent such behavior. An effective station location/alignment optimization model should contain an integrated part to address such competitions while not sacrificing too much computation efficiency.

As discussed above, ridership attraction is the most popular design criterion used in locating rail transit stations. It significantly affects the cost and benefit of urban rail transit systems: more riders generally mean higher operation cost, but also greater benefit in congestion reduction, energy saving, pollution reduction, safety improvement, etc. Researchers have shown consistent concerns about the accuracy of ridership forecasts for urban rail transit systems [Flyvbjerg, et al., 2005; Balaker and Kim, 2006].

Existing station location models represent ridership attraction either by the number of rail transit users calculated with simple mode choice models, or by the alignment coverage estimated as line coverage or station coverage [Laporte et al., 2000]. Such methods are quite simplified compared to the transit ridership forecasting models that are used in rail transit planning studies. The next section will briefly review these ridership forecast models and practices.

2.2.2. Transit Ridership Forecasting Models

Transit ridership forecasting models are mathematical models that can predict the future usage of transit systems, based on the land use pattern, transportation infrastructures and traveler's behavior of the study area. Groundbreaking work on transit ridership forecasting goes back more than 30 years. Boyle [2006] in his review on transit ridership forecasting classified the existing studies into two general categories, namely those based on the traditional four-step travel demand forecasting approach and those using regression models to explain the change in ridership with a group of variables such as the demographic characteristics within the transit covered area, quality/fare of rail transit service, and travel conditions by competing modes. Some other approaches not belonging to these two groups also appeared in the literature. Thus the rest of this section presents existing ridership forecast approaches in three general categories.

Four-Step Transit Ridership Forecasting Models

Considering the transit network as an integral part of the urban transportation system, the four-step based transit ridership forecasting models predict transit ridership via the following four-step procedure [Vuchic, 2005]:

- Trip generation determines the number of trips generated in each origin and attracted to each destination.

- Trip distribution assigns the total trips from each region to different destinations.

- Modal split predicts the percentage of trips that would use transit system.

–   Trip assignment estimates the number of trips taking each transit route.

Each of these four steps can employ different mathematical models. For example, trip generation has used multivariate regression models or cross-classification analysis [Meyer and Miller, 2001]. For trip distribution we can choose from a growth factor model, gravity model, or intervening opportunity model. For modal split we may use either aggregate models based on multivariate regression [Marshall and Grady, 2006] or cross-classification, or disaggregate models predicting individual behaviors based on utility theory. Trip assignment can be all-or-nothing, user equilibrium or system optimal assignment.

There are also different variations of the traditional four-step procedure. Examples include both combined models, which can solve two or more steps concurrently, and feedback modeling processes, which can solve different steps iteratively.

Several metropolitan planning organizations (MPOs) and other regional agencies have developed transit ridership forecasting tools based on the four-step travel models. The North Central Texas Council of Governments integrated a transit analysis process within its four-step regional travel demand model using the commercial software package TransCAD [NCTCOG, 2007]. The post-distribution mode choice function was realized with nested logit and multinomial logit models for different trip types. The Knoxville Regional Transportation Planning Organization employed a pre-distribution mode choice module and a convergence based feedback loop from traffic assignment to trip distribution in its regional travel demand model

[Conger, 2007]. Also using the feedback loop from traffic assignment to trip distribution, the Atlanta Regional Commission applied the post-distribution mode choice [Rousseau, 2007].

The fundamental problem of the traditional four-step based transit ridership forecasting models, in addition to costly and time-consuming nature, is that these models are region-level models and they are not calibrated at a resolution or with the kinds of variables necessary to conduct a fine-grained analysis for detailed transit analysis [Walters and Cervero, 2003]. In most cases, especially in suburban and exurban settings, the regions (usually traffic analysis zones) are too large to capture the characteristics of the neighborhoods surrounding transit stations.

Regression Based Transit Ridership Forecasting Models

Unlike the four-step procedure, regression-based transit ridership forecasting models, or sometimes called direct ridership forecasting models, consider the transit network more as an independent system. These models directly estimate the transit demand by examining the environmental, system and behavioral characteristics associated with transit ridership [Taylor and Fink, 2003]. Based on the study area, regression based transit ridership forecasting models fall into three major groups, namely region-level models, route-level models, and station-level models.

A region-level direct demand model estimates the total transit demand generated in a region or between regions, where the region is usually a traffic analysis zone or metropolitan statistical area. Examples of the in-region ridership estimation models include the work of Chatterjee, et al. [2002], Taylor, et al. [2004], Zhao

[2005] and Thompson and Brown [2006]. Stratifying transit ridership by trip type or time of day, these studies used either one-equation model that uses transit service variables as exogenous variables, or two-equation models to account for the potential interrelations between transit ridership and service decisions. For the inter-region ridership estimation, Thompson [1997] proposed an intra-suburban transit travel demand model calibrated with travel survey data from Sacramento, California. The model could estimate the potential transit trips among different census tracts based on their potential of producing/attracting transit trips and the difficulty of using transit between tracts.

In a pioneering study on route-level ridership estimation, Pushkarev et al. [1982] found strong explanatory power in demographic and transportation variables (such as downtown size, population density, geographic population distribution, auto ownership, and radial line length) on the number of passengers attracted by single transit corridors to central business districts. Later route-level multivariate regression studies include a model introducing transit service quality as explanatory variables [Kemp, 1981] and models designed for different periods of days [Stopher, 1992; Hartgen and Horner, 1997]. Two other models were proposed on a more refined route-segment level [Peng, et al., 1997; Kimpel et al., 2000; Kimpel, 2001]. The earlier model incorporated transit demand, supply and inter-route effects in a simultaneous regression system to estimate transit demand by fare zone, time of day and direction. The latter model included two steps, which estimated transit service reliability first and then estimated the transit demand accordingly.

The aforementioned region- or route-level direct demand models seem relatively accurate in their numerical studies. However, the potential for applying these models for detailed transit analysis might be limited, as they usually adopt the assumption of evenly distributed characteristics in the study region or along the transit route, which is often too insensitive to changes in residential or employment patterns around transit stations. This partly explains the appearance of various station-level transit ridership estimation models in recent years.

Cervero and Zupan [1996] applied multi-variable regression to develop two station-based ridership models, one for commuter rail and the other for light rail. Both models, though groundbreaking, still leave much opportunity for improvement on technical grounds, such as including more explanatory variables for better model fits and eliminating data uncertainties for estimation reliability.

Walters and Cervero [2003], in their study of the BART system, developed ridership models with better fit by establishing statistical relationships between station boarding/alighting and the characteristics of transit services and surrounding neighborhoods, such as station-area population and employment within walking radius, catchment-area population, feeder bus service level, parking supplies, train frequency and train vehicle type.

Chu [2004] developed a station-level ridership model as a part of the TLOS (transit level of service) program for the Florida Department of Transportation. The model related average weekday boarding at a transit stop with six categories of factors, namely socio-demographics in a catchment area, TLOS value, street

environment for pedestrians, accessibility to population and employment, interaction with other modes and competition with other transit stops.

Saur et al. [2004] developed a quick-response approach for rail passenger forecasts. Their approach also used multivariate regression to examine the effect of station-level variables, including surrounding land use and service characteristics at a given station, on the ridership of three different rail services (i.e., heavy rail, light rail, and commuter rail).

Lane et al. [2006] presented two improved multi-variable regression models to estimate daily station boardings by taking into account reverse commute trips to employment areas outside CBD and by introducing service-related variables such as travel speed, fare, and midday headways.

Station-based direct transit ridership forecasting models consistently validate the reliability of various demographic and transportation variables in predicting rail ridership with reasonable accuracy. However, there are also limitations to these models.

- Most models experience problems with multicollinearity, or a high level of correlation, between explanatory variables.
- Some models find transit service-related variables as significant explanatory variables for predicting transit ridership. This correlation may only arise from the fact that service supply is usually determined by transit demand. In design practice, increasing transit service may prove ineffective for increasing transit use.

- Most models do not distinguish between direct and transfer boardings and therefore cannot quantify trip-linking and provide a means of analyzing the effects of transfer opportunities on ridership.

- Most station-level ridership forecasting models cannot generate trip tables with origin-destination information or even guarantee a balanced demand, as boarding/alighting are usually estimated for each station independently.

- Differences in settlement patterns and travel behaviors may erode the relevance of a model's structure and therefore limit its generalizability among regions or over time.

Other Methods

There are less complex alternatives for regional modeling in practice. The simplest level is pivot point or elasticity analysis, which is mostly used for short-term service changes such as frequency and fare changes. This approach uses current ridership of the target system and service elasticities calculated from similar service changes as the foundation for estimating future ridership. Similar methods can also be applied to estimate the ridership increase resulting from an expanded service area of the transit network. However, this approach has a number of disadvantages. For example, the generic elasticities may be too inaccurate to substitute for local values; the approach is unable to respond to differences in the residential or employment patterns along routes.

Reference class forecasting has also been proposed as an alternative to conventional modeling. Flyvbjerg et al. [2005], based on their observation of inaccurate demand forecast in world-wide transportation infrastructure projects, suggested that the most effective way to improve forecasting accuracy is probably not improved models but more realistic assumptions and systematic use of empirically based assessment of uncertainty and risk. Accordingly, the authors proposed the reference class forecasting method, which took a so-called "outside view" on the project under study. The outside view, established on the basis of information from a relevant reference class of past projects, placed the project in a statistical distribution of outcomes from these reference projects.

Kikuchi and Millkovic [2001] used hierarchical fuzzy inference to predict transit ridership at individual stops. This approach is similar to the traditional cross-classification approach to trip generation modeling, with the boundaries of the discrete classes being fuzzy. However, this fuzzy rule–based model has shown little advantage over traditional regression-based methods, as the sensitivity of predictions to changes in continuous patronage-influencing factors is limited by grouping them in discrete categories.

In summary, the prediction of transit ridership has attracted a great amount of intellectual attention. Various methodologies over different scales of study areas have been proposed and examined with empirical transit data. They usually showed promising results with relatively accurate ridership predictions, whereas when examining real-world transit projects researchers still found significant prediction

errors [Flyvbjerg, 2005; Balaker and Kim, 2006]. Further efforts are justified to bridge this gap between academic studies and empirical work. One more observation is that the existing transit ridership prediction models are all behavioral models based on trend extrapolation, rather than normative models for achieving rational goals. On the other hand, most transit projects have some normative nature, i.e., these projects are expected to help shift the land use pattern or even urban form to a more desirable way in the long run. Ways of accounting for this normative nature in transit ridership forecasting still need further study.

2.3 Cost and Benefit Associated with Urban Rail Transit

2.3.1 Cost of Urban Rail Transit

This section investigates major cost items associated with the operation of urban rail transit systems. Such an investigation is essential since these costs are the most common criteria employed in evaluating alternative urban rail transit designs. To better present the wide range of costs, this section summarizes and reviews these cost items in the following two major categories:

− Those for providing the rail transit service; and

− Those for utilizing the rail transit service.

2.3.1.1. Cost Items for Providing Rail Transit Service

A number of studies [Hay, 1982; Vuchic, 2005] have discussed the costs for providing urban rail transit services. These studies generally classify the costs into

several groups. Construction Costs, or Capital Costs, are those costs spent to build and equip urban rail transit systems and are mostly one-time investments. Operation and Maintenance (O&M) Costs are those costs spent to run urban rail transit systems, which are generally incurred continuously over the entire life cycle. A third group, Environmental Cost, reflects damage to the surrounding environment. In most cases, the environmental cost is not an out-of-pocket cost and is difficult to represent with monetary values.

The rest of this section discusses the cost items of construction costs, operation and maintenance costs as well as environmental costs in more detail, along with brief reviews of existing cost estimation practice.

Construction Costs

Construction costs cover all cost items to build the urban rail transit system, which includes acquiring right-of-way, performing earthwork, laying down tracks, constructing structures/stations, purchasing vehicles, and paying for labor, energy, and/or various other miscellaneous items (e.g., fence and guardrails, underground utilities, drainage). A July 1977 revision of the Uniform System of Accounts for Railroads prescribes a series of capital accounts numbered 1 to 77, entitled "Road and Equipment" [Hay, 1982].

Baum-Snow and Kahn [2005] examined the 16 major metropolitan areas that established or expanded rail transit infrastructure from 1970 to 2000. They estimated that federal, state, and local governments spent more than $25 billion on construction.

These construction costs depended on a variety of factors, such as right-of-way category, horizontal/vertical alignment, station complexity and local conditions.

Transit ROW is the strip of land on which a transit line operates. The alignment classification system recommended by Vuchic [2005] included three basic alignment classes: *exclusive* ROW fully separates with surface traffic and can be at-grade, on embankment, on aerial viaduct, in a cut or in tunnel; *semi-exclusive* ROW has longitudinal separation along the alignment but with at-grade crossings for vehicles and pedestrians; and *nonexclusive* ROW has trains operating in a shared space with motor vehicles, other transit vehicles, or pedestrians. Generally, exclusive right-of-way requires the highest investments, but provides the fastest and most reliable service.

Among the horizontal alignment elements, the length of the line is the main factor affecting the overall construction costs. According to their relations with horizontal distance, construction costs can be briefly classified into three groups [Hay, 1982]. Certain cost items vary directly with the length of the rail line, such as tracks, ties, ballast, and the labor for their construction or application. Others have little or no relation to the horizontal length and include major yard and terminal facilities. A third class of semi-variable costs occur when the increase in length requires additional facilities or where unusual local factors exist. Examples include right-of-way, bridges, tunnels, road crossings, and etc.

Compared with the horizontal alignment, the vertical alignment usually has a much greater impact on the construction costs. The vertical profile will basically determine the amount of earthwork and the need for special structures (i.e., tunnels or

bridges). In very general terms, the lowest investment is required for ground-level rail transit, particularly if purchased right-of-way is minimized when the proposed rail transit line can be built along existing corridors of railways or freeway medians.

The investment cost of a rail transit station depends on its size, complexity and local conditions. One of these costs is associated with automobile parking, which consumes a very large area (20-30 m square per space). Investment cost per parking space in surface lots is $3000-6000 [Vuchic, 2005]. Garages require less area, but the investment cost is as high as $20,000 per space.

In practical projects, the planners generally prepare the capital cost estimate by breaking the defined alignment into logical geographical limits or line segments to establish quantities such as length of track, item counts, pipe lengths etc. Historical data are then used to set up the unit cost and lump sum cost items [MARTA, 2007].

Operation Costs

Unlike capital expenditures that occur only at the time of purchase or construction, the operation and maintenance costs occur throughout the life of the urban rail transit system. Railway operation expenses have been classified by the ICC into four general account categories, i.e., maintenance of way and structure, maintenance of equipment, transportation-rail line and general and administrative [Hay, 1982]. These expenses include not only the actual costs of moving trains and handling passengers, which are usually called direct costs, but also those indirect costs that, although necessary to the operation, cannot be directly assigned to the movement of any particular train or passenger.

Station distribution also affects the rail operation cost. For equal local conditions, the incremental cost per station is constant when stations are far apart, but decreases slightly when spacing becomes so short that trains cannot reach maximum speed, as the incremental time and energy consumption per stop are then slightly reduced. [Vuchic, 2005]

As an example estimation approach in practical projects, a MARTA study [2007] prepared the operation and maintenance cost estimates with the cost estimating models calibrated based on MARTA budget experience. The input data for the models were estimates of future operating statistics, equilibrated using the forecasted transit ridership. Similarly, McBrayer [2003] estimated the operation and maintenance cost for a hypothetical LRT route by assuming the operation statistics such as train capacity and service strategies.

Environmental Costs

Construction and operation of a new urban rail transit system may also significantly affect the entire environment. Such environmental impacts are often considered as the most important issues in the modern rail transit system construction projects. As part of the environmental review process required under the National Environmental Policy Act of 1969, a study of the impacts that would result from new rail transit projects must be taken into consideration of the design process. Such impacts may include several aspects.

First, the rail transit line may impact environmentally sensitive areas or disrupt human activities in the existing land-use system located along the alignments

or near the proposed stations. The former may include areas such as parks, wetlands, and historic/archaeological sites, whereas the latter include the residences, businesses, community facilities, churches, etc.

Secondly, the rail transit system generates air pollution during its operation. The total emissions depend on the vehicle miles traveled, the average speeds on network as well as the potential emissions from the investment itself [MARTA, 2007]. The noise impacts are sometimes represented with the number of households (residential houses and apartment buildings) within 200 feet of the transit lines [MARTA, 2007].

Finally, urban rail transit operations may interfere with surface traffic and cause additional delays for those vehicles, especially if priority is given to rail transit at at-grade crossings. Cline [1986] examined the delays that could be attributed to LRT at-grade crossings with NETSIM. The study found that the volume to capacity (v/c) ratio was the major factor in the delay and most of the effects were localized near crossings. Chandler and Hoel [2004] also examined the effects of light rail crossings on average delays experienced by vehicles with VISSIM. They examined four scenarios for the effects of variable traffic volumes and light rail crossing frequencies. The case studies found that the average additional delays from light rail transit crossings increase with increasing light rail crossing frequencies and increasing traffic volumes up to the roadway's capacity. As the road reaches an oversaturated condition, the average total delays continue to increase, but the difference in total delays with and without LRT decreases from the unsaturated condition. A report by the Institute of Transportation Engineers [ITE, 1992] examined

the design and operations of LRT at-grade crossings and supported those previously stated volumes corresponding to guidelines for LRT grade separation, namely 15,000-20,000 vehicles per day for acceptable grade crossings and 20,000-40,000 for possible grade crossings.

2.3.1.2. Costs for Utilizing Rail Transit Service

The costs for utilizing rail transit service are also called the user costs, which include the travel time costs, transit fare, and other potential access costs such as parking and fuel costs if passengers choose to drive to/from rail transit stations. Transit fare is the price paid by the public for using the transit service and these payments constitute the operating revenues [Hay, 1982]. Thus, if the planners consider an urban rail transit system as a single entity consisting of both service providers and service users, transit fare will become an internal cash flow (or "transfer payment") and need not appear in the planning process. Other user costs are sensitive to alignment as well as location of stations where users can board or alight, and thus should be considered in the proposed optimization problem.

Travel Time Costs

The travel time cost can be computed with users' value of time generated externally and their estimated travel time, which generally consist of the following parts [Vuchic, 2005]:

    − Access time: travel time to station or from station to destination;

- Waiting time: time between passenger arrival at a station and the time of train departure. For frequent transit service that closely adheres to its schedule the average waiting time is approximately half of the headway. For longer headways (usually >6min), passengers begin to use time tables and adjust their arrivals, so that the average waiting time becomes somewhat shorter than for random arrivals and remains approximately constant for longer headways. [Bowman and Turnquist, 1981]

- In vehicle time: the travel time between the boarding station and the alighting station. This time may include running time and station standing (or dwell time). Given exclusive right-of-way, the running time depends mainly on the technical characteristics of the track-vehicle system as well as the geometric features of the rail alignment [Hay, 1982]. Nonexclusive right-of-way may introduce additional delays due to the interference by other traffic, such as potential intersection delay time at grade crossings [Vuchic, 2005]

- Transfer time: time needed for passengers to switch between different transit lines at a station

Other Access Costs

When walking is the only mode that passengers use to access the transit system, travel time can be considered as the only major user cost, as in most downtown areas. However, transit users may incur other access costs if they use other traffic modes to access transit stations.

Bus is one of most widely used options for transit access. The time for riding buses to stations has been counted in the access time, and the bus fares are excluded from user cost.

Park-and-Ride, as another popular access option, has been in use since the 1930s. Travelers drive to transit stations via uncongested local roads and then use rail transit to avoid the heavily congested surface corridors. In this case, the user cost should take into account the fuel and emissions associated with the access travel, and the parking costs if the park-and-ride facility is operated by a private firm.

In summary, Table 2.1 itemizes the major costs that should be considered in the design of an urban rail transit system. These costs cover not only items for providing rail transit services, but also those for utilizing transit services.

Table 2.1 Classification of Urban Rail Transit Costs

Classification		Examples
Construction Costs		Right-of-way, Earthwork, Structures
Operation and Management Costs		Train/station operation, Maintenance
Environmental Costs		Wetland disturbance, Noise, Emission
User Costs	Travel Time Costs	Access time, Waiting time, In-vehicle time
	Other Access Costs	Parking, Travel costs to/from stations

2.3.2 Benefit of Urban Rail Transit Operations

This section investigates the other group of critical justification criteria for urban rail transit projects, i.e., the potential benefits urban rail transit systems could bring. The economist defines user benefit as being equivalent to the value which

travelers expect to receive from making trips, as measured by the maximum amount which travelers would be willing to pay for those trips [Wohl and Hendrickson, 1985]. However, in most transit related studies, benefits are usually specified in terms of cost savings as travelers switch to the transit system from other modes. The discussion in this section will fall into the latter category.

Various studies have discussed such benefits, among which the most obvious ones are, for transit users, the savings of travel cost they would otherwise pay to complete the trips without transit services, and, for service providers, the direct operation revenue from transit fare. As mentioned in the previous section, the latter can be treated as an internal cash flow and thus left out of the planning process.

A broad range of other urban rail transit benefits are also mentioned in the literature. In the short term, urban rail transit systems, once put in use, can provide an alternative travel option and attract travelers from surface road networks. Such a ridership shift may directly result in congestion relief, environmental benefits and safety improvements. These benefits are often among the most important justifications for building an urban rail transit system. In the long run, urban rail transit systems are expected to function as a catalyst in introducing more accessible transit supportive land use, reducing automobile ownership and further helping the ridership shift from auto to more efficient transit mode [Hess and Ong, 2001; Podobnik, 2002; Switzer, 2002; Renne, 2005; Frank, et al., 2006]. Moreover, an urban rail transit system may also improve the overall mobility and equity by providing service for people who cannot afford cars or who cannot drive cars.

Nelson et al. [2007] used a regional transport model to estimate the benefits of the local transit system to transit users and the congestion-reduction benefits to motorists. They estimated that weekday rail produces about $833 million in traveler benefits and a welfare benefit of $5.16 per rail trip. Litman [2006a] reported that U.S. rail transit services produce economic benefits $19.4 billion in annual congestion cost savings, $8.0 billion in roadway cost savings, $12.1 billion in parking cost savings, $22.6 billion in consumer cost savings, and $5.6 billion in traffic accident cost savings. The rest of this section will present a review of these benefits in more detail.

Congestion Relief

An urban rail transit system, due to its high capacity, is considered by Zaretsky [1994] to be a much more efficient way to move people around a metropolitan area than automobiles. Thus, ridership shift from auto to rail transit is expected to reduce surface road congestion and congestion-associated delay, vehicle operating costs, emissions and stress, especially in the absence of road pricing [Lewis and Bekka, 2000].

Although the percentage of transit shift from surface travel, or more importantly, auto trips, varies a lot in practice [FTA, 2002; Litman, 2006a; Hilton, 1976; Lave, 1998; Richmond, 2001], experience has shown that transit does help to reduce traffic congestion. Garrett [2004] found that traffic congestion growth rates declined in several U.S. cities after the operation of light rail systems. In Baltimore the congestion index increased at an average rate of 2.8% annually before light rail,

but only 1.5% afterward. In Sacramento the index increased 2.2% annually after light rail service, compared with a 4.5% growth rate before.

Such congestion relief occurs mostly because rail transit offers an alternative on the most congested corridors. As urban traffic tends to maintain equilibrium, grade-separated transit of relatively high travel speeds acts as a pressure-relief value to reduce the equilibrium congestion level on these roadways [Litman, 2006b]. Various studies have indeed found that door-to-door travel times for motorists tend to converge with those of grade-separated transit [Mogridge, 1990; Lewis and Williams, 1999; Vuchic, 1999].

Energy and Emission Reductions

Shapiro et al. [2002] found substantial environmental gains based on empirical data. For every passenger mile traveled by Americans in 1998, rail travel consumes about a third of the energy of private automobiles, SUVs and light trucks, due to its high mechanical efficiency and load factors. Similarly for every passenger mile traveled by Americans in 1999, rail transit produces less than 10 percent as much carbon monoxide and volatile organic compounds, and little more than half as much carbon dioxide. They concluded that greater use of public transportation offers the most effective strategy available for achieving significant energy savings and environmental gains without imposing new taxes, government mandates or regulations.

By providing surface traffic congestion relief, rail transit provides even greater energy and emission reduction benefits.

Safety Improvement

Traffic accidents are among the largest causes of deaths and disabilities for people nowadays, imposing billions of dollars in economic losses annually. Rail transit has been shown to improve traffic safety in several studies. Gleave [2005] found that in the U.K., deaths and injury rates on established urban rail transit systems are quite low compared with those on other modes. Litman [2006a] also showed that rail transit cities have significantly lower per capita traffic death rates in the US. Another study by Kenworthy and Laube [2000] used international data to demonstrate that per capita traffic fatalities decline with increased transit ridership.

User Cost Savings

Rail transit users avoid the travel cost they would otherwise pay to complete the trips without transit services. This will bring cost savings potentially for several reasons.

First, the impact of surface road congestion is little for rail transit with semi-exclusive and exclusive right-of-way, which leads to lower travel time, especially for those congested commuter corridors during peak hours. Travel time savings have been found in several empirical studies. For example, Lewis and Bekka [2000] calculated the travel time index for Washington, D.C.'s I-270 corridor and demonstrated a 4 million hour saving of delay in 1999. Litman [2006a] also found that per capita congestion delay is significantly lower in cities with high quality rail transit systems than in otherwise comparable cities with little or no rail service. Even if there are no time savings, costs per hour of using high quality transit service might

be lower than for driving, as transit allows passengers to relax and work [Litman, 2006b].

Another part of cost savings attribute to parking, as both parking cost and parking search time tend to be lower around stations near residence areas than in CBD.

Finally, for some equity-justified service to disadvantaged people, who often cannot drive, transit substitutes alternatives that have to include the cost of a driver (such as taxi service or chauffeured automobile travel by family members and friends).

Other Benefits

Except for the aforementioned direct benefits, urban rail transit systems can also bring other benefits, such as

- Economic development benefits: Public transit can result in various economic development benefits to local areas. Increased property value along rail transit lines and around rail transit stations due to improved accessibility and livability is one of the most widely discussed economic development benefits [Cockerill and Stanley, 2002; Eppli and Tu, 2000; Garrett, 2004; Lewis and Bekka, 2000; Smith and Gihring, 2006; Weinstein and Clower, 2002]. Another economic development benefit is the increased localized economic activities, employment, and income [Litman, 2006a; Miller et al., 1999] The high land use density and clustering associated with rail transit can also reduce the costs of providing public services and

increase productivity due to improved accessibility and network effects [Haughwout, 2000; Litman, 2003].

- Equity: Generally, people who are more inclined to use transit are those physically, economically and socially disadvantaged compared with drivers, such as those who do not own automobiles, those whose income levels put them below the poverty level and elderly people over the age of 65. The operation of urban rail transit improves the mobility and accessibility for these people, and thus increases equity among all the travelers.

- Option value: Transit services provide option value, referring to the value people place on having a service available even if they do not currently use it [TRB, 2002]. For example, drivers may like to have the choice of using transit services in case of personal and community-wide emergencies, such as when a personal vehicle has a mechanical failure, or a disaster limits automobile travel.

Sections 2.3.1 and 2.3.2 provide a detailed review of major costs and benefits that should be considered in the design of urban rail transit systems. Generally, higher investment should yield lower operating cost and/or higher service quality, thus attracting more passengers and introducing more benefits. Several different methodologies have been employed to examine these trade-offs and to perform a thorough design evaluation.

The most popular evaluation approach is the economic evaluation of the investment and its effectiveness, such as benefit-cost analysis or computation of rate of return. The cost is the life cycle cost, which accounts for the initial capital outlay in

terms of the useful lives of the individual capital cost components [Hay, 1982]. In other words, the capital cost needs to be considered on an equal basis along with the annual operation and maintenance cost. Economic evaluation of designs provides a single, easily understood measurement for each alternative design, but, for many transit projects, it is either inadequate or misleading as many impacts are difficult to measure in monetary units. Thus, the economic evaluation approach is more appropriate when unquantifiable factors are limited, such as when the alternative designs differ mostly in technical characteristics but serve the same function and have similar system performance [MARTA, 2007].

Another evaluation approach is to directly evaluate the system performance, which has been defined as being comprised of two elements, efficiency and effectiveness [Fielding and Glauthier, 1976]. Efficiency reflects production and is a measure of the ratio of service outputs to resource inputs, while effectiveness reflects consumption and is a measure of how well goals are met by the provision of service. Although mostly used for evaluation of existing systems, some measures, such as the cost per consumed output in Compin's study [1999], can also apply to the evaluation of design alternatives.

To accommodate those impacts that are hard to quantify, MARTA [2007] introduced a numerical scoring system in the evaluation process. Each of the performance measures was assigned a weight based on the importance of its corresponding system goals, and also a score from 1 to 10 to reflect how well the alternative performed. The overall measure of each alternative is a composite score weighted over all criteria.

2.4 Summary

In summary, this chapter has provided a comprehensive review of those existing research efforts in three aspects of urban rail transit designs, i.e., determination of track alignment, selection of transit stations and evaluation of design alternatives. These are essential design issues that determine the infrastructure of an urban rail transit system and impact its performance once in use.

Although the review has reported various theoretical studies and practical procedures in each of the three aspects, there still exist great needs to overcome existing technical deficiencies and to further improve the theoretical study for real-world applications.

Need for methodology to integrally design track alignment and station locations

The existing literature on optimal design of rail transit systems has two distinct groups. Alignment optimization models are only constrained by end terminals, whereas station location models generally assume given alignments or simply not consider the alignment related geometry constraints.

This tendency of theoretical work overlooks the complex interaction between the two components in real-world design practices. On one hand, planners sometimes predefine a few major stations other than the two end terminals for the alignment to follow. Thus the alignment shall be constrained by the geometric requirements at stations. On the other hand, stations must allow feasible alignments, and thus station location models should to some extend account for the geometry requirements of alignments.

Lai and Schonfeld [2010, 2012] presented a methodology that can effectively incorporate the aforementioned interactions between alignment design and station location selection, which is covered in greater detail in this dissertation. The methodology is able to generate feasible alignment connecting the two end terminals and/or intermediate stations.

Need for evaluation framework for comprehensively comparing design alternatives

Interestingly, a similar diversifying tendency appears in optimal design models of rail transit systems when it comes to the evaluation of different design alternatives. Existing alignment optimization models, focused more on the cost side, mostly try to minimize the overall system cost. On the other hand, station location models pay more attention to the benefits that rail transit systems can bring, and mostly try to maximize the potential ridership shift from other transportation modes.

In reality, however, planners have to carefully consider both the cost and the benefit aspects in designing a rail transit system. Higher cost generally associates with more competitive service quality, and potentially greater cost savings. Yet the extra cost saving may or may not justify the additional investment. In response, this dissertation develops an evaluation framework that can account for the tradeoffs between cost and benefit, and thus more comprehensively evaluate different design alternatives.

Existing optimal design models of rail transit systems have left out some potentially important decisions in real-world design practices. An example is the type of stations to be proposed. Most station location studies simply predefine station types as pedestrian-based and/or Park-and-Ride, by assuming fixed station cost and given station attraction radius/area.

Station types, however, have impacts on a variety of aspects of an urban rail transit system, such as land needs, construction cost, user access mode, and potential attraction radius. A station without Park-and-Ride facility generally needs smaller sites and costs less to construct and operate, but has a limited attraction area and attracts fewer passengers.

Generally, densely populated urban areas or suburban centers have a significant portion of transit riders that access transit services on foot so as to avoid the costs associated with owning/driving/parking a vehicle, while suburban areas typically have many transit riders to access by auto. However, a clear cut criterion for deciding station types is lacking. This dissertation incorporates such decision variables in the modeling process, and lets the optimization procedures automatically select the type for each station.

Need for representing some critical constraints in the analytical formulations

Existing optimal design models of rail transit systems have also left out some critical constraints in real-world design practices. An example is the geometric

requirements at stations. Both alignment design and station location studies have overlooked these requirements and treated stations as individual points.

However, a rail transit station has its own layout and thus enforces some geometric constraints on the passing track alignment. The platform alongside rail tracks from which passengers board or alight from trains needs to be located along a tangent track of a minimum length with maximal longitudinal slope for safety and cost considerations. It makes more sense to embed such constraints directly in the station location/ alignment optimization process, rather than shift the stations later. This dissertation incorporates such constraints in the optimization procedures.

Need for excluding unnecessary or unrealistic assumptions in the design process

Most existing optimal design models have employed unrealistic assumptions to simplify some essential relations in rail transit systems. For example, when calculating travel time along the rail transit line, most models assume the availability of average speed or direct travel time data. Some models directly enforce predefined minimal station spacing constraints to guarantee an acceptable operation speed.

In practice, the speed and thus the travel time is determined by vehicle dynamics, which should account for both the horizontal/vertical alignment and the station spacing. Generally, tighter horizontal curves and larger gradients should be avoided, in order to achieve higher operation speed. Station locations on the vertical profile also affect the operation efficiency due to the need for stop, acceleration and deceleration at stations. This dissertation thus explicitly captures such vehicle

dynamics in the modeling process and examines their impacts on the design of urban

rail transit systems.

Chapter 3: Optimizing Rail Transit Alignments

That Connect Several Major Stations

The literature review in Chapter 2 indicates neglect of the complex interactions between station locations and track alignment: alignment optimization models are only constrained by end terminals, whereas station location models generally assume given alignments or simply ignore the alignment related geometry constraints. In real-world design practices, however, planners often predefine a few major stations between the two end terminals at major demand points and/or transfer centers. The geometry requirements at these intermediate stations have to be taken into account in designing a realistic alignment. This is also a crucial issue in developing integrated optimization models that can concurrently select transit station locations and optimize track alignment between stations.

This chapter proposes a practical rail transit alignment optimization methodology, which can generate alignments that pass through preset station locations while meeting the special geometry constraints at these stations. Section 3.1 presents the model formulation, elaborating the decision variables and various geometry constraints to account for in the design problem. Section 3.2 proposes a heuristic based on a Genetic Algorithm to efficiently search for solutions while interacting with the supporting GIS system. This section details how the algorithm represents a practical route alignment, incorporates the design constraints and computes the optimal solutions. The current objective function minimizes

construction cost but the search algorithm is designed to optimize any function that can be evaluated with available GIS data.

3.1 Model Formulation

3.1.1 Components of Rail Transit Alignment

This thesis models a 3-dimensional rail transit alignment with two separate components: the horizontal alignment defines physically where the track goes (the XY plane), while the vertical alignment defines the elevation along the horizontal alignment (the Z component).

The horizontal alignment consists of a series of tangents joined with circular curves and spiral transition curves, as illustrated in Figure 3.1. Spiral transition curves are used here to accommodate the changes in curve radius and to provide safer and more comfortable passenger conditions. The key points of the horizontal alignment include:

- PI_i: the hypothetical point of intersection for two adjacent tangent tracks.

- TS_i: the point of change from tangent to spiral pertaining to PI_i

- SC_i: the point of change from spiral to circle pertaining to PI_i

- CS_i: the point of change from circle to spiral pertaining to PI_i

- ST_i: the point of change from spiral to tangent pertaining to PI_i
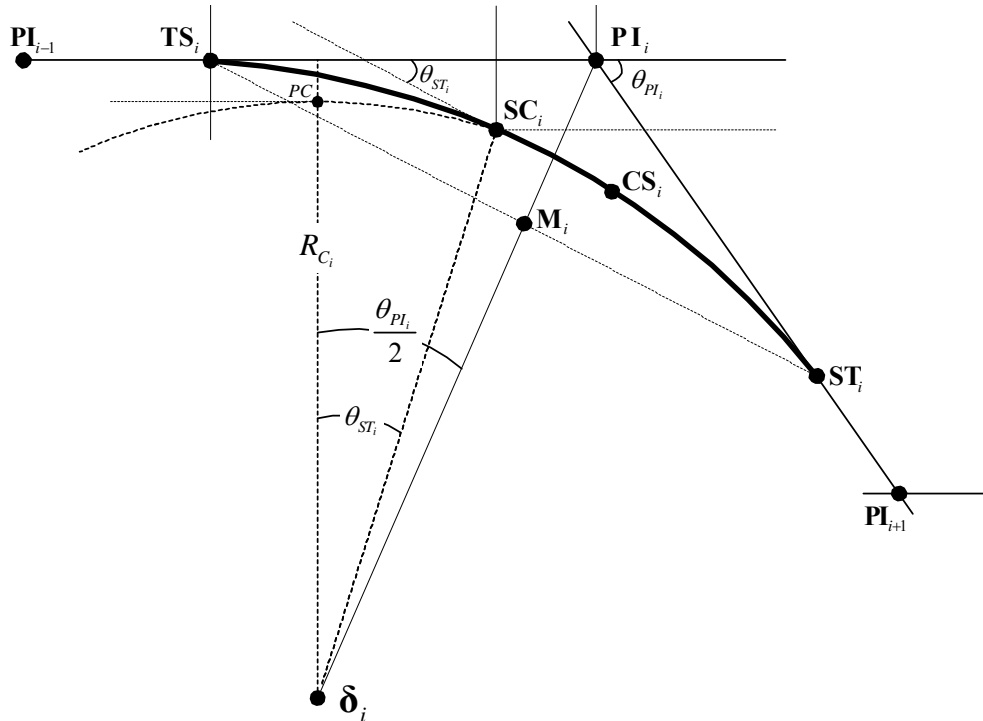
Figure 3.1 Illustration of Horizontal Alignment Components [Kang, 2008]

The key geometry variables for a horizontal alignment are defined below:

δ_i = The central point of the curved section between SC_i and CS_i

R_{C_i} = The radius of the circular curve between SC_i and CS_i (ft)

l_{C_i} = The length of the circular curve between SC_i and CS_i (ft)

θ_{PI_i} = Deflection angle at PI_i (radians)

M_i = The middle point of the line segment connecting TS_i to ST_i

l_{ST_i} = Total length of spiral curve from TS_i to SC_i (ft)

θ_{ST_i} = Central angle of spiral arc l_{ST_i}, called "spiral angle" (radians)

x_{ST_i} = Total tangent distance from TS_i to SC_i with reference to initial tangent (ft)

y_{ST_i} = Total tangent offset at SC_i with reference to TS_i and initial tangent (ft)

p_{S_i} = Offset from the initial tangent to the point of curvature of the shifted circle (ft)

k_{S_i} = Abscissa of the shifted point of curvature to TS_i (ft)

L_{TS_i} = Tangent distance from TS_i to PI_i (ft)

The vertical alignment consists of a series of grades joined by vertical curves, as illustrated in Figure 3.2. The key points of the vertical alignment include

- VPI_i: the hypothetical vertical point of intersection for two adjacent grades.

- VPC_{i+1}: the point of change from tangent to vertical curve pertaining to PI_i

- VPT_i: the point of change from vertical curve to grade pertaining to PI_i
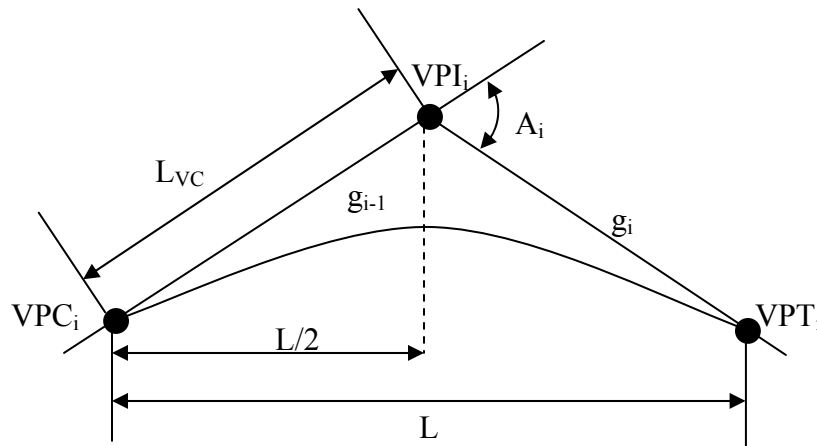
Figure 3.2 Illustration of Vertical Alignment Components

The key geometry variables for a vertical alignment are defined below:

$g_i =$ The gradient of the tangent connecting VPI_i and VPI_{i+1} (%)

$A_i =$ The algebraic difference in gradients between g_{i-1} and g_i (%)

$L_i =$ The total length of vertical curve pertaining to VPI_i (ft)

$L_{VCi} =$ The tangent distance from VPC_i to VPI_i (ft)

$L_{VTi} =$ The tangent distance from VPT_{i-1} to VPC_i (ft)

3.1.2 Geometric Constraints for Rail Transit Alignment

To generate rail transit alignments that meet various geometry requirements derived from the engineering practice, this chapter incorporates three groups of geometric constraints.

3.1.2.1 Special Alignment Constraints at Rail Transit Stations

Existing optimization models on locating rail transit stations typically assume the station is a single point on the alignment and enforces no additional geometry constraints. In reality, a rail transit station has its own layout and elements, for example, the platform alongside rail tracks from which passengers board or alight from trains. For safety and cost considerations, most station design practices require the platform to be located along a minimum length of tangent track whose grade and cross slope do not exceed given thresholds [Davies, 2007; FWTA, 2007; I-70 Coalition, 2008]. In response, this chapter introduces two groups of special alignment constraints for rail transit stations.

For horizontal alignment, the model requires that the entire length of station should be located on a tangent section and the tangent before and after the central point of the station must exceed a minimal length.

Similarly for vertical alignment, the model requires that the station be located on a tangent with a grade not exceeding the maximal allowable grade, and the tangents before and after the central point of the station exceed a minimal length.

It should be noted that some railway design guide [MTA 2007] also requires a desired minimum grade at passenger stations to ensure adequate track drainage. However, in engineering practice, drainage can be easily maintained with properly designed cross slopes of ballast, subballast or subgrade and the underdrain system. Thus, the proposed model does not incorporate such minimum grade requirements.

3.1.2.2 General Geometric Requirements for Horizontal Track Alignment

This group of requirements includes simple boundary values, including minimum tangent length between curves, minimum circular curves radius, minimum circular curves length and minimum spiral length. The group also covers those constraints related to track superelevation, which include maximum applied superelevation and unbalanced superelevation, spiral superelevation runoff constraint, spiral jerk rate constraint and spiral roll rate constraint.

Tangent Length between Curves

The minimum length of tangent track between curved sections is based on passenger comfort and vehicle truck/wheel forces.

Based on the AREMA Manual, The Desired Minimum Tangent Length $L_t = \max\{3V, 200\,ft\}$, where V is the operating speed in mph. This formula is based on vehicle travel of at least 2 seconds on the tangent track between two curves. This criterion has been used for various transit designs in the U.S.

Other studies employ an Absolute Minimum Tangent Length depending on the selected operation vehicles. The Maryland Transit Administration [MTA, 2007]

used 40 feet in the Purple Line segment from Bethesda to Silver Spring. The same value is recommended by Washington Metropolitan Area Transit Authority [WMATA, 1976]. The Transportation Research Board (TRB) track design handbook for Light Rail Transit [TRB, 2000] recommends an Absolute Minimum Tangent Length of 100 feet.

Circular Curve Radius

A Desired Minimum Circular Curve Radius is recommended since track maintenance and wheel squeal is drastically increased on curves with a small radius [TRB 2000]. Both AREMA and TRB specify 500 feet for the desired minimum curve radius. WMATA specifies 920 feet for dedicated right-of-way and 285 feet in street running.

The Absolute Minimum Circular Curve Radius is determined by the characteristics of the railway vehicles. Both AREMA and TRB specify 82 feet absolute minimum curve radius. WMATA specifies 300 feet for dedicated right-of-way and 82 feet in street running.

Circular Curve Length

The minimum circular curve length is dictated by ride comfort and is not related to vehicle physical characteristics. The Absolute Minimum Length of a Superelevated Circular Curve should be 45 feet. The Desired Minimum Circular Curve Length is generally determined by $L = 3V$, where V is the design speed through the curve, in mph

For compound curves consisting of a starting spiral, a circular curve and an ending spiral, the length of the circular curve added to the sum of one-half the length of both spirals is an acceptable method of determining compliance with the above criteria.

Superelevation

Superelelevation serves to counteract the centrifugal force acting radically outward on the vehicle as it travels through the curve [TRB 2000].

Equilibrium superelevation is the amount of superelevation that would be required to make the resultant force from the center of gravity of the rail transit vehicle perpendicular to the plane of the two rails and halfway between them at a given speed. If a curved track is superelevated to achieve equilibrium at a given speed, a rail vehicle passenger would experience no centrifugal force through the curve at that speed. Equilibrium superelevation is usually determined by the following formula:

$$E = (E_A + E_U) = 3.96V^2/R \qquad\qquad (3.1)$$

where,

$E =$ Equilibrium superelevation, inches,

$E_A =$ Applied superelevation, inches,

$E_U =$ Unbalanced superelevation, inches,

$R =$ Radius of curve, feet,

$V =$ speed, mph,

In practice, full equilibrium superelevation E is rarely installed in track. This would require excessively long spiral transition curves. It could also produce passenger discomfort on a train that is moving much slower than the design speed or stopped in the middle of a steeply superelevated curve. Therefore, only a portion of the calculated equilibrium superelevation is commonly installed as applied superelevation E_A [TRB 2000]. Desired values of applied superelevation can be determined from the following formula:

$$E_A = 2.64V^2 / R - 0.66 \qquad\qquad (3.2)$$

Unbalanced superelevation is the difference between the equilibrium and applied superelevation [TRB, 2000]. The desired balance between applied superelevation and unbalanced superelevation shall be defined by the following relationship:

$$E_U - E_A / 2 = 1 \qquad\qquad (3.3)$$

As a guideline, TRB[2000] recommended the maximum values for applied and unbalanced superelevation as follows:

$E_A = $ 4 inches (desired), 6 inches (absolute)

$E_U = $ 3 inches (desired), 4.5 inches (absolute)

Transition Spirals

Spiral transition curves are used to gradually build into the superelevation of the track and limit lateral acceleration during the horizontal transition of the rail vehicle as it enters the curve. Various types of spirals found in railway alignment design include AREMA Ten Chord, PTC/SEPTA, Cubic, Bartlett, Hickerson clothoid,

and ATEA. For LRT design, it is recommended that spiral transition curves should be clothoid spirals. It is recommended that spirals should be used on all main line track horizontal curves with radius less than 10,000 feet wherever practicable.

Spiral curve length should be greater of the lengths determined from Runoff Rate, Jerk Rate and Roll Rate, and greater than the absolute minimum spiral length.

Superelevation runoff rate is defined as the allowable rate at which actual superelevation is introduced and removed along the given length of spiral. Therefore,

$$L_S = \frac{E_A}{runoff \quad rate} \tag{3.4}$$

AmTrak requires 31 feet per inch of superelevation, which gives:

$$L_S \geq 31E_A \tag{3.5}$$

The Jerk Rate (J) is defined as the rate of change of the lateral acceleration and is expressed in feet/second^3. From Chapter 11, Section 3.5.7.9 of AREMA, the equation is:

$$L_S = \frac{A_{lat}}{J} \times \frac{E_u}{E_{u,max}} \times 1.46V \tag{3.6}$$

where,

A_{lat} = Lateral acceleration (ft/second^2)

$E_{u,max}$ = Maximum unbalanced superelevation (inch)

Assuming a maximum lateral acceleration of $0.1g$, a jerk rate based on a passenger comfort level of 0.04g/second as recommended by WMATA [1976] and

the maximum unbalanced superelevation of 3 inches gives:

$$L_S \geq 1.22E_uV \qquad (3.7)$$

The vehicle roll rate is expressed as the rate of change of vehicle roll due to the effect of applied superelevation and is expressed in inch/second. The equation is:

$$L_S \geq \frac{E_a \times 1.46V}{roll\ \ rate} \qquad (3.8)$$

For the design LRT vehicle established for Purple Line system, this roll rate is limited to 1.56 inches/second. The resulting formula for minimum length of spiral is:

$$L_S = 0.94E_aV \qquad (3.9)$$

The absolute minimal spiral length is defined as 60 feet in TCRP report 57 [TRB, 2000] and 40 feet for WMATA [1976].

3.1.2.3 General Geometric Requirements for Vertical Track Alignment

This group of requirements enforces three boundary conditions: minimum vertical tangent length, maximum vertical tangent grade and minimum vertical curve length.

Tangent Length

The minimum desirable length of tangent between successive vertical curves is based on both passenger comfort level and vehicle suspension system wear. In TCRP

report 57 [TRB, 2000], the minimum length of tangent between vertical curves should be $3V$ or 100 feet, whichever is greater. The absolute minimum tangent length is required to be 40 feet in the same report, while it is 50 feet for WMATA [1976].

Tangent Grade

Maximum grades in track are controlled by vehicle braking and tractive efforts. On mainline track, civil drainage provisions also establish a minimum recommended profile grade. According to TCRP Report 57 [TRB, 2000], grades in the range of 0% to 4% are acceptable.

Vertical Curves

All changes in grade are connected by vertical curves. Vertical curves shall be provided at all points of vertical grade intersections where the algebraic difference between grades is greater than 0.15% (0.0015 ft./ft.). The length of vertical curve shall be determined as follows:

- o Desirable Minimum Length of Vertical Curve

  - WMATA [1976]: $100(g_i - g_{i+1})$

  - TRB [2000]: $200(g_i - g_{i+1})$

- o Absolute Minimum Length of Vertical Curve

  - WMATA [1976]: 50 feet

  - TRB [2000]: for crest curve $\dfrac{(g_i - g_{i+1})V^2}{25}$;

    for sag curve $\dfrac{(g_i - g_{i+1})V^2}{45}$

where g_i = forward grade of the vertical curve (%)

g_{i+1} = backward grade of the vertical curve (%)

V = vehicle speed (mph).

3.1.3 Cost Formulation for Rail Transit Alignment

The proposed model aims to minimize the total costs incurred in the construction of the rail transit system. As the problem addressed here assumes the station locations are preset, the station construction costs will be the same for all feasible alternatives. The objective function is thus the construction costs of tracks, which include right-of-way cost, earthwork cost, track installation cost, as well as structure cost for bridges and tunnels. The formulation of these costs is elaborated in the following Section 3.2.2.

3.2 Algorithm

The proposed method is designed to generate track alignments consistent with engineering practice, which are hard to model with simple mathematical functions. Besides, the cost function involves various non-linear or even discontinuous local conditions, and has a non-differentiable structure. Thus, this chapter presents a heuristic search method based on a Genetic Algorithm (GA) for efficiently solving this problem.

A GA is a search technique widely used to solve a variety of large-scale optimization problems. Inspired by evolutionary biology, GAs typically utilize a

computer simulation in which a population of candidate solutions evolves toward better solutions. The evolution starts from a population of completely random solutions and proceeds in iterations (generations). In each generation, the fitness of the population is evaluated, while multiple individuals are stochastically selected from the current population based on their fitness and modified with genetic operators to form a new population for the next generation [Goldberg, 1988].

The flow chart of the proposed heuristic, shown in Figure 3.3, follows classic GA procedures. The remaining section will detail two of the key steps, i.e., the representation of candidate track alignments and the calculation of the cost function.

Input Location of Preset Major Stations

Generate Candidate Alignments
Set Up Cutting Planes
Initial Intersection Points P_i — New Intersection Points P_i
Generate Horizontal Alignment
Generate Vertical Alignment

Fitness Function Evaluation
Earthwork Cost
Bridge/Tunnel Cost
Right-of-Way Cost
Track Cost

GIS Database

Genetic Operators

Meet Stop Criteria?

N

Y

End
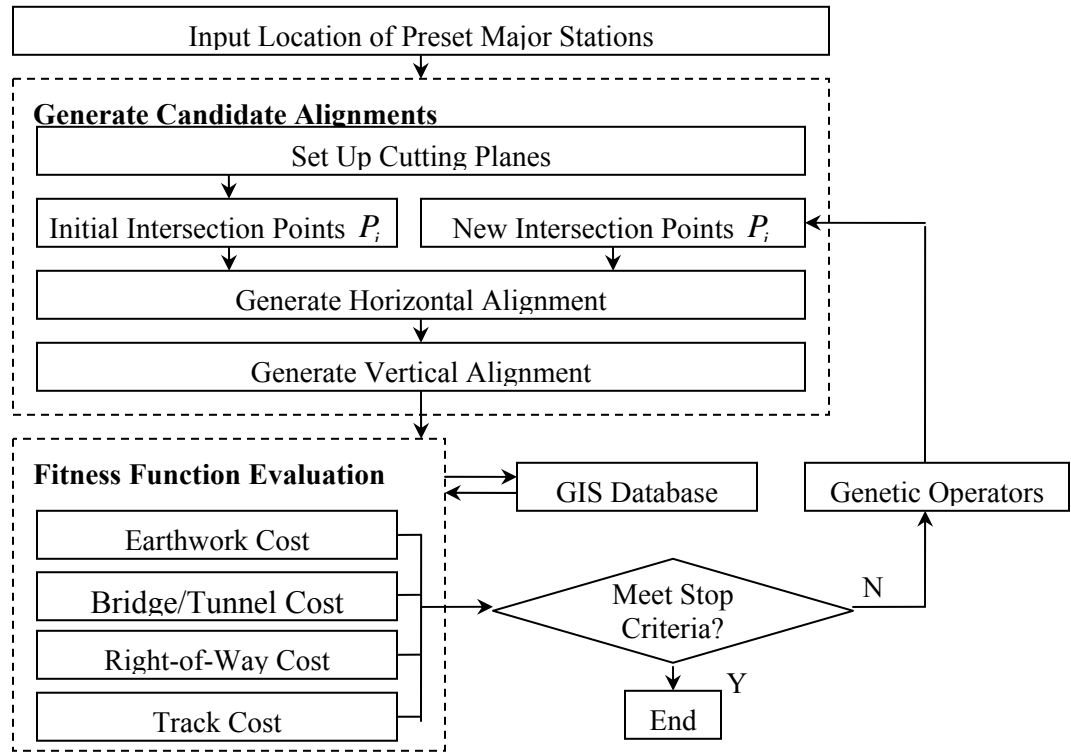
Figure 3.3 Flow Chart of the Proposed GA Heuristic

3.2.1 Representation of Candidate Track Alignments

To represent each candidate track alignment, the proposed heuristic employs the cutting plane concept [Jong, 1998] to define the PI/VPI, and then uses two numerical procedures to insert the appropriate curves.

3.2.1.1 Setting up Cutting Planes

As illustrated in Figure 3.4, cutting planes between two points are perpendicular to the straight line connecting these two points and to the X-Y plane. The forward and reverse tangents along the three dimensional alignment will intersect each cutting plane i at point $P_i(x_i, y_i, z_i)$, whose projection on the X-Y plane defines the PI_i for horizontal alignment. Its elevation also defines the elevation of VPI_i at the corresponding location. Here, O_i is the point along the straight line connecting the two end points where the ith cutting plane crosses, whereas d_i and Z_i are, respectively, the abscissa and ordinate of P_i on the ith cutting plane relative to O_i.
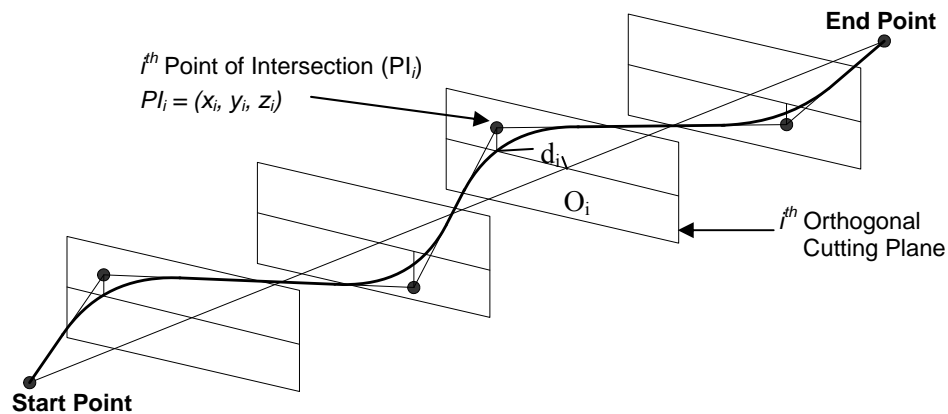
Figure 3.4 Illustration of Cutting Plane Concept

To precisely locate the major stations on a candidate alignment, the proposed heuristic compares two different strategies:

- Use equal spacing cutting planes between the two end terminals. Skip the cutting plane if it is too close to a major station to allow a minimal tangent before/after the station, as shown in Figure 3.5;
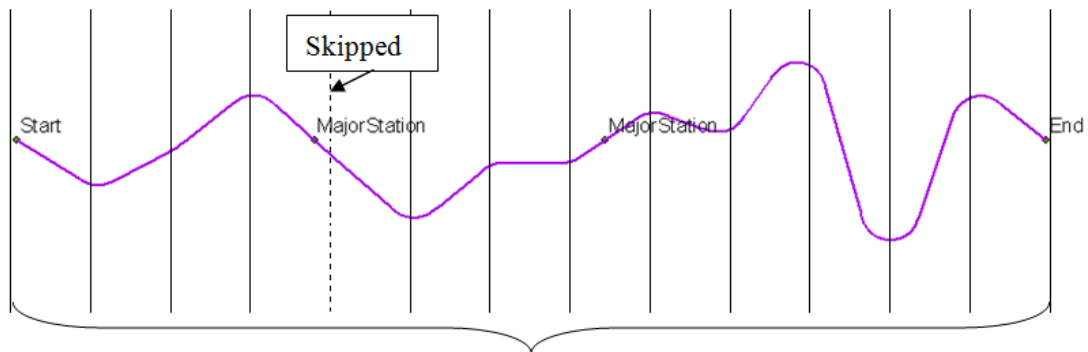
Figure 3.5 Equal Spacing Cutting Planes between End Terminals

- Use unequal spacing cutting planes between two neighboring stations, which allows for a minimal tangent before/after stations, as shown in Figure 3.6.
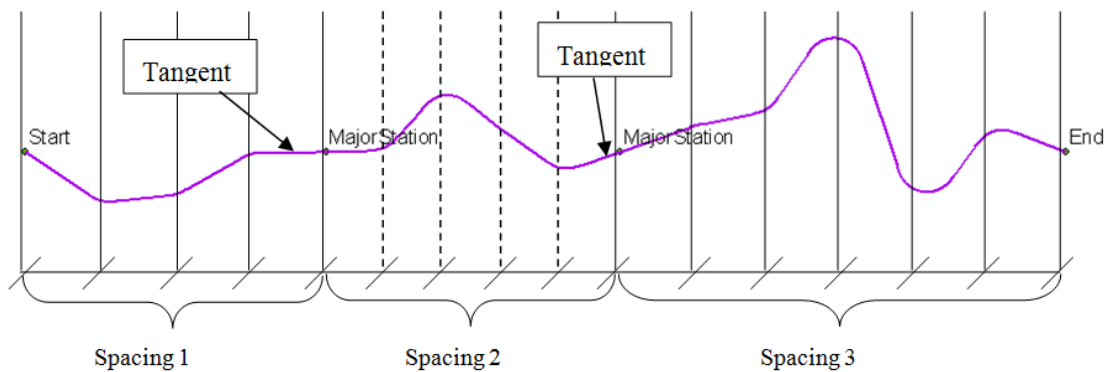
Figure 3.6 Unequal Spacing Cutting Planes between Stations

3.2.1.2 Locate Intersection Points

The proposed GA procedure aims to optimize the locations of the intersection points PI_i on the cutting planes. The distances from PI_i to the center of cutting planes O_i are assumed to follow a uniform distribution within a preset range centered on O_i when generating the initial solutions or new solutions in successive generations. Thus, the algorithm will tend to produce a z-shape alignment if the centers of the cutting planes are along the single straight line connecting the two end terminals, as in previous studies. To overcome this limitation and generate smoother alignments, this heuristic also introduces a successive center point generation method, as shown in Figure 3.7. The center point of the ith cutting plane O_i is now defined as the intersection points between cutting plane i and the straight line connecting PI_{i-1} and the end terminal, which is generated successively in the algorithm for cutting planes $i = 1,\ldots,n$.
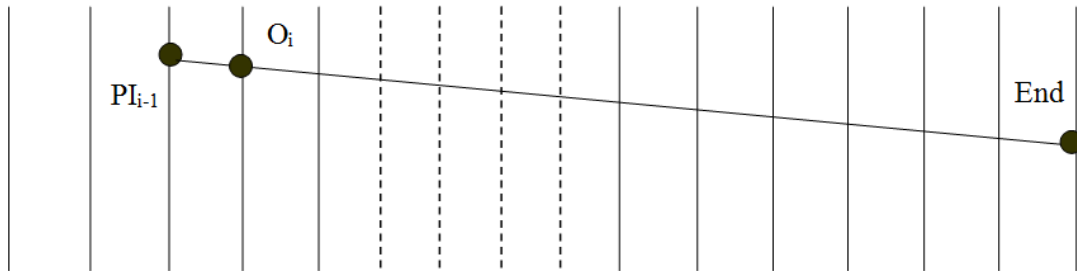
Figure 3.7 Successive Center Point Generation Method

After the PI_i s are either located randomly or evolved from the previous generation, the proposed heuristic uses two numerical procedures to generate the horizontal alignment and vertical alignment.

3.2.1.3 Generate Horizontal Alignment

Step 0: Calculate intersection points $PI_i, i = 1, \ldots, n$. Then for each cutting plane i

Step 1: Initialization

Calculate intersection angle:

$$\theta_{PIi} = \cos^{-1} \frac{(PI_i - PI_{i-1}) \cdot (PI_{i+1} - PI_i)}{\|PI_i - PI_{i-1}\| \cdot \|PI_{i+1} - PI_i\|} \tag{3.10}$$

Calculate the distance between PI_{i+1} and PI_i:

$$D_i = \|PI_{i+1} - PI_i\| \tag{3.11}$$

Set $\quad TS_i = PI_i, \ ST_i = PI_i, \ l_{C_i} = 0, \ l_{ST_i} = 0, \ R_{C_i} = R_{C,\min} \tag{3.12}$

Step 2: Find the curve radius, where V = design speed of the rail transit vehicles, $E_{A,\max}$ = maximum applied superelevation, K_{runoff} = $1/$ superelevation runoff rate, K_{roll} = $1.47 /$ vehicle roll rate, and $K_{Jerk} = 0.0488 /$ jerk rate.

$$R_{Ci} = MAX(R_{C,\min}, \quad R_1, \quad R_{2i}, \quad R_{3i}, \quad R_{4i}, \quad R_{5i})$$

$$R_1 = \frac{2.64 V^2}{(E_{A,\max} + 0.66)}$$

$$R_{2i} = \frac{l_{C,\min} + l_{ST,\min}}{\theta_{PIi}}$$

$$R_{3i} = \frac{(l_{C,\min} - 0.66 K_{runoff}) + \sqrt{(l_{C,\min} - 0.66 K_{runoff})^2 + 4\theta_{PIi} \cdot 2.64 K_{runoff} V^2}}{2\theta_{PIi}} \tag{3.13}$$

$$R_{4i} = \frac{(l_{C,\min} - 0.66 K_{roll} V) + \sqrt{(l_{C,\min} - 0.66 K_{roll} V)^2 + 4\theta_{PIi} \cdot 2.64 K_{roll} V^3}}{2\theta_{PIi}}$$

$$R_{5i} = \frac{(l_{C,\min} + 0.67 K_{Jerk} V) + \sqrt{(l_{C,\min} + 0.67 K_{Jerk} V)^2 + 4\theta_{PIi} \cdot 1.32 K_{Jerk} V^3}}{2\theta_{PIi}}$$

Step 3: Find the superelevation and spiral length, where E_{Ai} and E_{Ui} are respectively the applied superelevation and unbalanced superelevation at cutting plane i

$$E_{Ai} = 2.64V^2 / R_{Ci} - 0.66$$

$$E_{Ui} = 1 + E_{Ai} / 2$$

$$l_{STi} = MAX(l_{ST,min}, \ l_{1i}, \ l_{2i}, \ l_{3i}), where$$

$$l_{1i} = K_{runoff} \cdot E_{Ai}, \ l_{2i} = K_{Jerk} \cdot E_{Ui} \cdot V, \ l_{3i} = K_{roll} \cdot E_{Ai} \cdot V$$

$$(3.14)$$

Step 4: Find TS_i and ST_i

$$\theta_{STi} = l_{STi} / (2R_{Ci})$$

$$l_{Ci} = R_{Ci} \times (\theta_{PIi} - 2\theta_{STi})$$

$$x_{STi} = l_{STi} \times [1 - \theta_{STi}^2 / 10 + \theta_{STi}^4 / 216 + \theta_{STi}^6 / 9360 + \theta_{STi}^8 / 685440]$$

$$y_{STi} = l_{STi} \times [\theta_{STi} / 3 - \theta_{STi}^3 / 42 + \theta_{STi}^5 / 1320 - \theta_{STi}^7 / 75600 + \theta_{STi}^9 / 6894720]$$

$$p_{Si} = y_{STi} - R_{Ci} \times [1 - \cos(\theta_{STi})]$$

$$k_{Si} = x_{STi} - R_{Ci} \times \sin(\theta_{STi})$$

$$L_{TSi} = k_{Si} + (R_{Ci} + p_{Si}) \times \tan(\theta_{PIi} / 2)$$

$$L_{Ti} = D_{i-1} - L_{TSi} - L_{TS\,i-1}$$

$$TS_i = PI_i + L_{TSi} \times (PI_{i-1} - PI_i) / \|PI_{i-1} - PI_i\|$$

$$ST_i = PI_i + L_{TSi} \times (PI_{i+1} - PI_i) / \|PI_{i+1} - PI_i\|$$

$$(3.15)$$

Step 5: Find M_i and δ_i

$$M_i = 0.5 \times [TS_i + ST_i]$$

$$\delta_i = PI_i + [(R_{Ci} + p_{Si}) \times \sec(\theta_{PIi} / 2)] \times (M_i - PI_i) / \|M_i - PI_i\|$$

$$(3.16)$$

Step 6: Find SC_i and CS_i

$$SC_i = R_{Ci} \times (M_i - \delta_i) / \|M_i - \delta_i\| \times R(\frac{\theta_{PIi}}{2} - \theta_{STi})$$

$$CS_i = R_{Ci} \times (M_i - \delta_i) / \|M_i - \delta_i\| \times R(-\frac{\theta_{PIi}}{2} + \theta_{STi}) \qquad (3.17)$$

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Step 7: Check feasible horizontal alignment criteria. The alignment is feasible only if the following conditions are met:

$$l_{Ci} \geq l_{C,min}, \quad L_{TSi} \leq D_i, \quad L_{TSi} \leq D_{i-1}, \quad L_{Ti} \geq L_{T,min} \qquad (3.18)$$

3.2.1.4 Generate Vertical Alignment

Step 0: Calculate vertical intersection points $VPI_i, i = 1, \ldots, n$. Then for each cutting plane i

Step 1: Adjust VPI_i elevation to satisfy maximum grade constraints, where z_i = elevation of VPI_i and h_i = distance from the starting station to VPI_i along the horizontal alignment.

$$g_i = 100 \times (z_{i+1} - z_i) / (h_{i+1} - h_i)$$

$$\text{if } g_i > g_{max} \Rightarrow \quad z_{i+1} = z_i + g_{max} \times (h_{i+1} - h_i) / 100, \quad g_i = g_{max} \qquad (3.19)$$

$$\text{if } g_i < -g_{max} \Rightarrow \quad z_{i+1} = z_i - g_{max} \times (h_{i+1} - h_i) / 100, \quad g_i = -g_{max}$$

Step 2: Find the vertical curve length, where K_{VC} is the minimal length needed per 1% change in grade.

$$A_i = g_i - g_{i-1}$$

$$L_i = MAX(L_{min}, K_{VC} \times A_i)$$

(3.20)

Step 3: Find VPC_i and VPT_i

$$L_{VCi} = \frac{L_i / 2}{Cos(0.5 \times \tan^{-1}(A_i / 100))}$$

$$VPC_i = VPI_i + L_{VCi} \times (VPI_{i-1} - VPI_i) / \|VPI_{i-1} - VPI_i\|$$

(3.21)

$$VPT_i = VPI_i + L_{VCi} \times (VPI_{i+1} - VPI_i) / \|VPI_{i+1} - VPI_i\|$$

Step 4: Check the feasibility of vertical alignment criteria. The vertical alignment is feasible only if the following conditions are met:

$$L_{VTi} \geq L_{VT,min}, where \quad L_{VTi} = \|VPI_{i-1} - VPI_i\| - L_{VCi} - L_{VCi-1}$$

(3.22)

3.2.2 Calculation of the Cost Function

As the construction cost of rail transit tracks involves very complex local conditions, such as topography features and private properties, a GIS-based program is developed in this chapter to effectively interact with the existing database for cost calculation.

3.2.2.1 Identify Cut/Fill Sections and Bridges/Tunnel Sections

Following engineering practice, the proposed program first applies the typical track cross sections at an equal spacing L_{CS} along the horizontal alignment on the corresponding elevation from the vertical alignment. The following data are then extracted from the GIS database: Z_i is the elevation from the vertical alignment at

cross section i; G_i is the ground elevation at cross section i along the horizontal alignment.

The program then uses these data to identify the cut/fill sections for generating earthwork costs and bridge/tunnel sections for generating structure costs. Here $\Delta_B =$ threshold of elevation difference at which a bridge becomes preferable to a fill section, $\Delta_T =$ threshold of elevation difference at which a tunnel becomes preferable to a cut section, $\Delta_{BC} =$ bridge clearance height, $N_{B,\min} =$ minimum number of consecutive bridge sections, and $N_{T,\min} =$ minimum number of consecutive bridge sections

- A bridge is constructed from section N_1 to N_2 if:

$$Z_i - G_i \geq \Delta_B \ \text{ or } \ (G_i = 0 \ \text{ and } \ Z_i \geq \Delta_{BC}), \qquad \forall i = N_1, N_1 + 1, \ldots, N_2$$
$$N_2 - N_1 \geq N_{B,\min}$$
(3.23)

- A tunnel is constructed from section N_1 to N_2 if:

$$Z_i - G_i \leq -\Delta_T \ \text{ and } G_i > 0, \qquad \forall i = N_1, N_1 + 1, \ldots, N_2$$
$$N_2 - N_1 \geq N_{T,\min}$$
(3.24)

- All other sections will be cut-and-fill sections.

3.2.2.2 Calculate Earthwork Cost for Cut/Fill Sections

Figure 3.8 illustrates a typical cut/fill section, where the gray line indicates the existing ground and the black line indicates the proposed ground. The program stratifies each cut/fill section with very small intervals, as shown with the dashed lines. The cut volume $E_{C,i}$ and fill volume $E_{F,i}$ are then calculated numerically for

82

each cross section i, based on the cut area $A^j_{C,i}$ and fill area $A^j_{F,i}$ between the proposed and the existing ground for each stratum j.

$$E_{C,i} = 0.5 \times (\sum_j A^j_{C,i} + \sum_j A^j_{C,i+1}) \times L_{CS}$$

$$E_{F,i} = 0.5 \times (\sum_j A^j_{F,i} + \sum_j A^j_{F,i+1}) \times L_{CS}$$

(3.25)
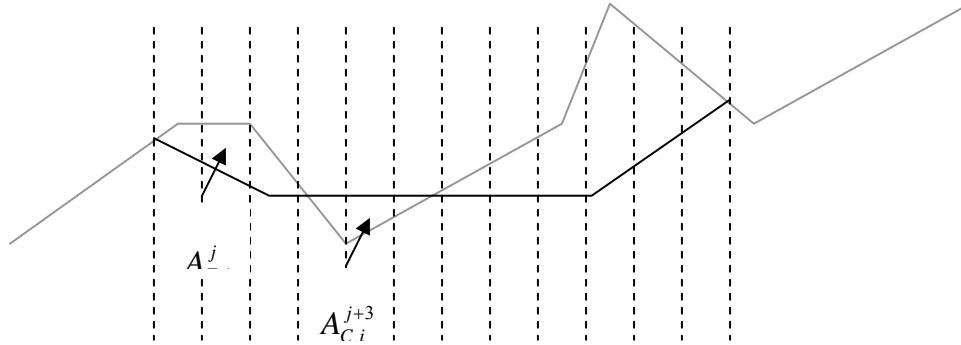
Figure 3.8 Earthwork of a Typical Cut/Fill Section

The total earthwork cost is calculated with the following equation, where E_N = the net earthwork, C_E = the total earthwork cost, s_e= earth shrinkage factor, K_C= unit cutting cost, K_F =unit filling cost, and K_l and k_b are unit transportation costs for, respectively, moving earth to a landfill and from a borrow pit:

$$E_N = \sum_i E_{C,i} s_e - \sum_i E_{F,i}$$

$$C_E = K_C \sum_i E_{C,i} + K_F \sum_i E_{F,i} + K_l \max(E_N, 0) - k_b \min(E_N, 0)$$

(3.26)

3.2.2.3 Calculate Structure Costs for Bridge and Tunnel Sections

For each bridge i, the proposed program uses an enumeration method to find the optimal span length L_{Bi} that minimizes the sum of superstructure C_{Bi}^U and substructure costs C_{Bi}^L [Jha et al., 2006]. The cost calculation is based on the predefined bridge width, the bridge length identified in the first step, and the pier height that depends on the vertical alignment and the ground elevation extracted from the GIS database

The cost for each tunnel i is much simpler here, and depends only on the predefined unit cost for tunnel excavation, the area of tunnel cross sections and the tunnel length.

3.2.2.4 Calculate Right-of-Way Costs

To calculate the right-of-way cost, the program first generates the right-of-way band along the horizontal alignment by connecting the edge points of each cross section. For a cut/fill section, the edge points are obtained by moving the outside tie-in points of the proposed cross section to existing ground with a buffer width. For bridges, the edge points are outside the bridge width by a buffer width. The tunnel cross section requires no right-of-way.

The program extracts data from Maryland Department of Planning MdPropertyView GIS database to locate all properties impacted by the right-of-way band, as shown in Figure 3.9. The right-of-way cost is the sum of values of these properties.
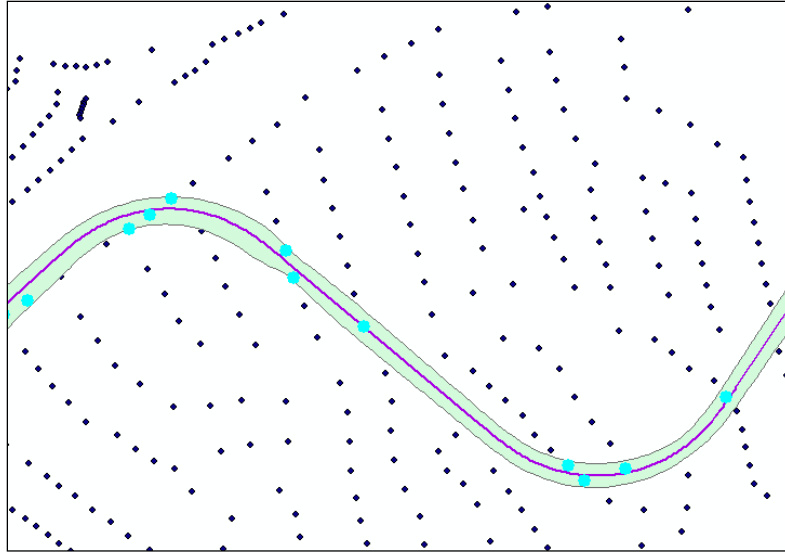
Figure 3.9 Calculation of Right-of-Way Cost

The total rail transit track construction cost is the sum of the earthwork cost, the bridge and tunnel cost, the right-of-way cost, and the track cost that depends only on the track length and a unit track installation cost.

3.3 Summary

This chapter proposes a practical rail transit alignment optimization methodology, which aims to generate alignments that pass through preset station locations. This is a major issue in the real-world design practice, where planners often predefine a few major stations at major demand points and/or transfer centers and the alignment has to accommodate the geometric requirements at these intermediate stations. This is also a crucial step towards the development of an integrated optimization models that can concurrently select transit station locations and optimize track alignment between stations.

With the proposed heuristic based on a Genetic Algorithm, the methodology can efficiently search among practically feasible alignments to minimize the construction costs. The heuristic employs the Points of Intersection at predefined cutting planes as decision variables. It then generates the alignments through a special procedure to satisfy three groups of geometry constraints, including the general geometric requirements for horizontal track alignment, the general geometric requirements for the vertical track alignment, and the special alignment constraints at rail transit stations.

The next chapter will use the Baltimore Red Line as a case study to illustrate the development of a computer program that integrates the proposed algorithm with a supporting Geographical Information System. Using this program, an extensive numerical study will be conducted to demonstrate the efficiency of the proposed optimization methodology in regions with complex topographical features. An extensive sensitivity analysis will also be included to provide some insights into the design problem.

Chapter 4: Case Study and Sensitivity Analysis

The previous chapter presented a practical rail transit alignment optimization methodology, which aims to help engineers design track alignment connecting several major stations. The methodology generates alignments that pass through preset station locations while meeting the special geometry constraints at these stations. The chapter also proposes a heuristic based on a Genetic Algorithm to efficiently search for solutions that minimize the overall construction cost.

To demonstrate the effectiveness of the proposed methodology and also to provide some insights into the design problem, Chapter 4 presents an extensive case study using a section of the Baltimore Red Line as an example. The section covers a 7-mile east-west transit corridor connecting five major stations in west Baltimore suburban residential areas, shopping areas and office parks.

Section 4.1 describes a computer program that integrates the proposed optimization heuristic with a background GIS database and user-friendly interfaces. Both the system framework and key modules are introduced. Section 4.2 presents elaborated numerical results when applying the program to the Baltimore Red Line Study. The results demonstrate that the proposed methodology can find very good solutions in regions with complex topographical features. A sensitivity analysis is presented in Section 4.3 to demonstrate the impacts of different design parameters and critical optimization parameters on the efficiency and effectiveness of the proposed methodology. As the sensitivity analysis demonstrates that cutting plane spacing may greatly impact the efficiency and effectiveness of the solution algorithm,

Section 4.4 presents an algorithm improvement which incorporates cutting plane spacing as decision variables in the optimization.

4.1 System Development

4.1.1 System Framework
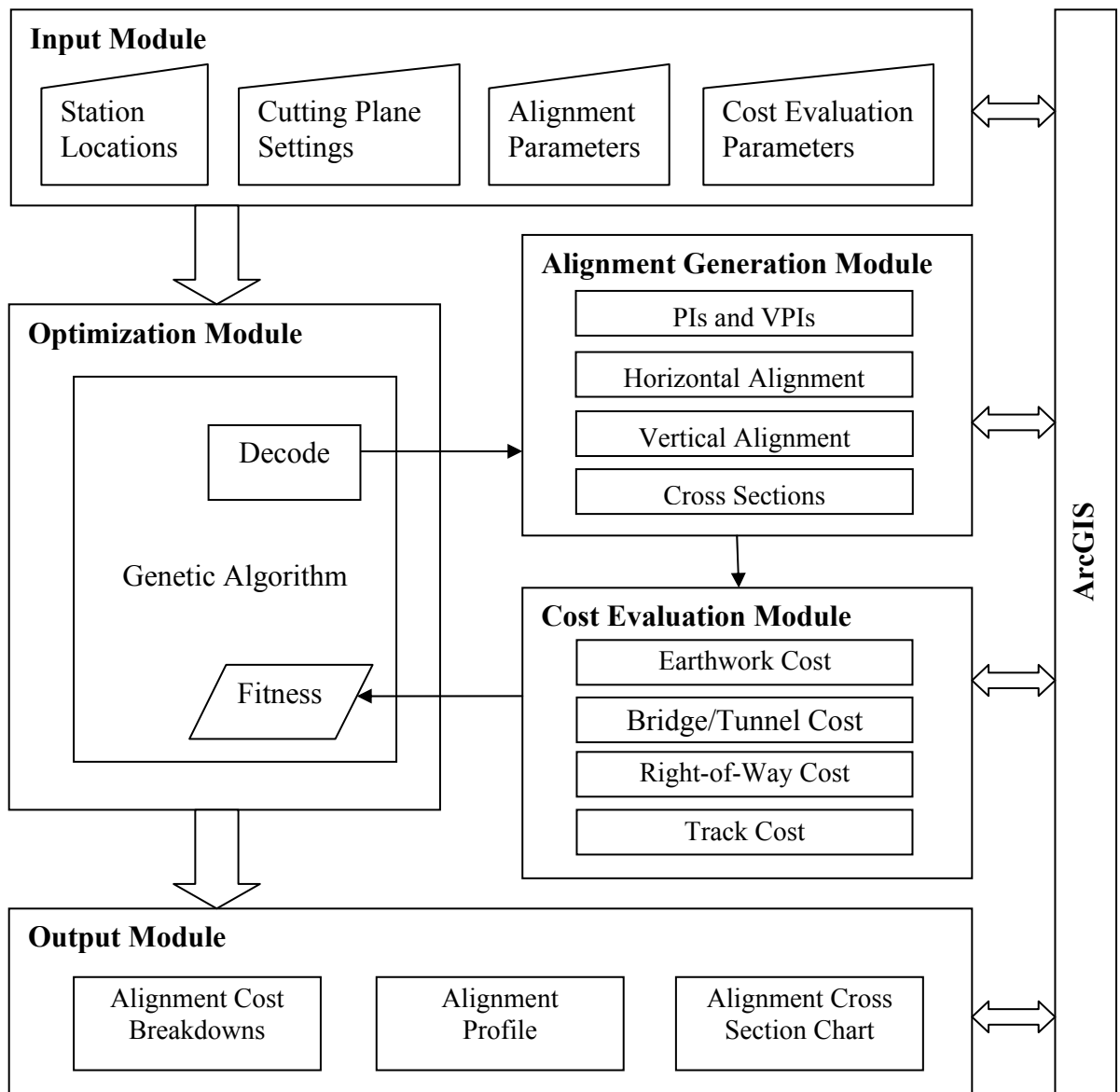
Figure 4.1 System Framework

Figure 4.1 presents the system framework of the proposed computer program for optimizing a rail transit alignment which connects several major stations. The system is programmed in Visual Basic and integrated with the ESRI ArcMap 9.2 GIS.

The system consists of the following five principal components:

- Input Module: This is employed by users to define the basic alignment settings, station locations, geometry constrains parameters, and cost evaluation related parameters.

- Optimization Module: A genetic algorithm is coded in this module to automatically search for the optimized railway alignment connecting the major stations.

- Alignment Generation Module: This is designed to create horizontal and vertical alignments after receiving inputs from the Optimization Module. All alignment points are calculated using the algorithms presented in Chapter 3 to satisfy the geometry constraints. They are then plotted in ArcGIS.

- Cost Evaluation Module: This evaluates the alignment generated from the Alignment Generation Module and returns the estimated cost as the fitness function to the Optimization Module.

- Output Module: This displays the customized output of the optimized rail transit alignment, and assists system users in examining the properties of the optimized alignment.

The proposed system framework features its module-based structure and its integration with the GIS platform. All modules are integrated by exchanging data inside the ArcGIS environment. On the other hand, each module is relatively independent with respect to its input and output needs, which offers the flexibility for further model updates or system expansion.

4.1.2 Principal System Modules

4.1.2.1 Input Module

This module consists of three interfaces for potential system users to input and adjust various design parameters before applying the alignment optimization algorithm:

- Basic Settings

Users can use this interface to define the basic settings for the candidate alignments, which include the locations of major stations on the alignment, rail transit design speed, alignment searching boundary, cutting plane settings, and cross section settings. Figure 4.2 presents a snapshot of the input interface for basic settings.

Figure 4.2 Input Interface for Basic Settings

- Alignment Geometry Parameters

The design criteria of alignment geometry are usually defined by the system owner/operator before the alignment planning stage. These criteria depend on the type of the rail system to be designed, and must account for many factors, including passenger comfort, vehicle-operating envelope and track safety requirements. Figure 4.3 illustrates the interface for the users to specify alignment geometry requirements, including parameters for horizontal alignment, for vertical alignment, and at stations.

Figure 4.3 Input Interface for Alignment Geometry Parameters

- Cost Evaluation Parameters

Cost parameters are critical for the fitness evaluation in the proposed GA-based solution algorithm. This interface is designed for system users to input the major cost related parameters, including cut/fill earthwork unit cost, bridge superstructure and substructure cost, and tunnel earthwork unit cost.

4.1.2.2 Optimization Module

The Optimization Module is the core component of the proposed GA-based solution algorithm. It functions to generate the initial population of completely random solutions and to implement genetic operators for solution evolution from one generation to the next. Each solution contains a set of PI and VPI locations for use in the succeeding Alignment Generation Module.

The Optimization Module consists of the following key functions:

- Encode(): subroutine to encode phenotype (PI and VPI locations on cutting planes) into genotype parameters (binary strings that GA operators work on) based on the major station locations and cutting plane settings; For each cutting plane, there are two decision variables: horizontal and vertical distances of PI to the center of the cutting plane. The elevation of each station is also a decision variable. The total length of the binary strings is:

$$L_b = (2 \times n_p + n_s) \times n_g \qquad (4.1)$$

where n_p is the number of cutting planes

n_s is the number of stations

n_g is the number of genes for each variable

- Decode(): subroutine to decode genotype (binary strings) into phenotype parameters (PI and VPI locations on cutting planes) for use by the Alignment Generation Module;

- SelectParent(): parent selection operator by roulette wheel algorithm;

- Cross(): crossover operator to breed two offspring from two parents based on crossover probability;

- Mutate(): mutation operator to introduce random mutation in a genotype based on mutation probability;

- STDREP(): genetic operator to insert offspring into population, for steady-state reproduction;

- FF(): subroutine to compute the fitness function by invoking the Alignment Generation Module and the Cost Evaluation Module.

4.1.2.3 Alignment Generation Module

The Alignment Generation Module functions to generate horizontal and vertical alignments that satisfy the geometry constraints from the set of PI/VPI locations. This module consists of the following key functions:

- FindGroundElevation(): subroutine to find the ground elevation by interacting with the background GIS database;

- GeneratePIsFromGA(): subroutine to generate PIs and VPIs coordinates from their locations on cutting planes

- HorizontalAlignment(): subroutine to generate horizontal alignment based on the procedures in Section 3.2.1.3.

- VerticalAlignment(): subroutine to generate vertical alignment based on the procedures in Section 3.2.1.4.

- GenerateCS(): subroutine to generate cross sections along the alignment at a given interval.

4.1.2.4 Cost Evaluation Module

The Cost Evaluation Module functions to estimate the overall construction cost of a candidate alignment generated from the Alignment Generation Module. The estimated cost is then fed back to the Optimization Module for fitness evaluation. The Cost Evaluation Module consists of the following key functions:

- FindBridgeTunnel(): subroutine to locate the alignment sections where bridges or tunnels will be built;

- BridgeCost(): subroutine to compute the bridge construction cost;

- TunnelCost(): subroutine to compute the tunnel construction cost;

- CSCutNFill(): subroutine to compute earthwork for cut-and-fill sections;

- GetROWArea(): subroutine to define the alignment ROW impact area;

- ROW(): subroutine to connect to the property GIS database and extract ROW costs;

- COST(): subroutine to compute the total costs for the candidate alignment.

Figure 4.4 shows an example ROW impact area from the GetROWArea() subroutine where the candidate alignment has a bridge and a tunnel.
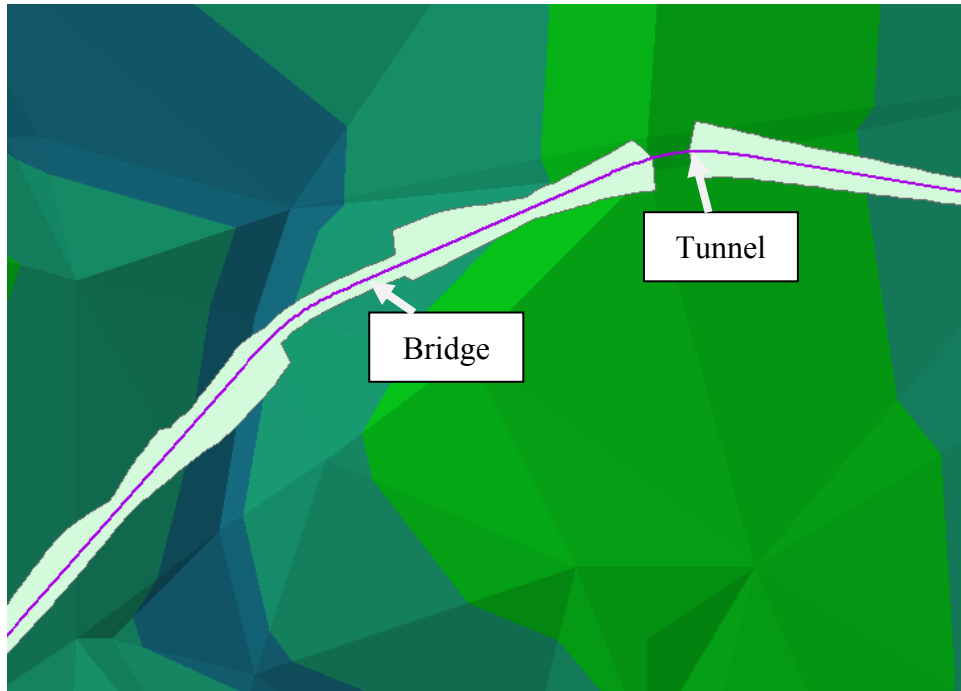
Figure 4.4 Example ROW Impact Areas

4.1.2.5 Output Module

Once the proposed alignment optimization heuristic reaches the predefined stopping criteria, the horizontal/vertical alignments is generated and displayed inside the ArcGIS environment. The Output Module can also provide the following information about the optimized alignment:

- Total costs and cost breakdowns, including earthwork costs, bridge costs, tunnel costs, ROW costs, and track costs.

- Alignment profile output, which compares the proposed top-of-rail profile with the existing ground elevation along the alignment, as shown in Figure 4.5.
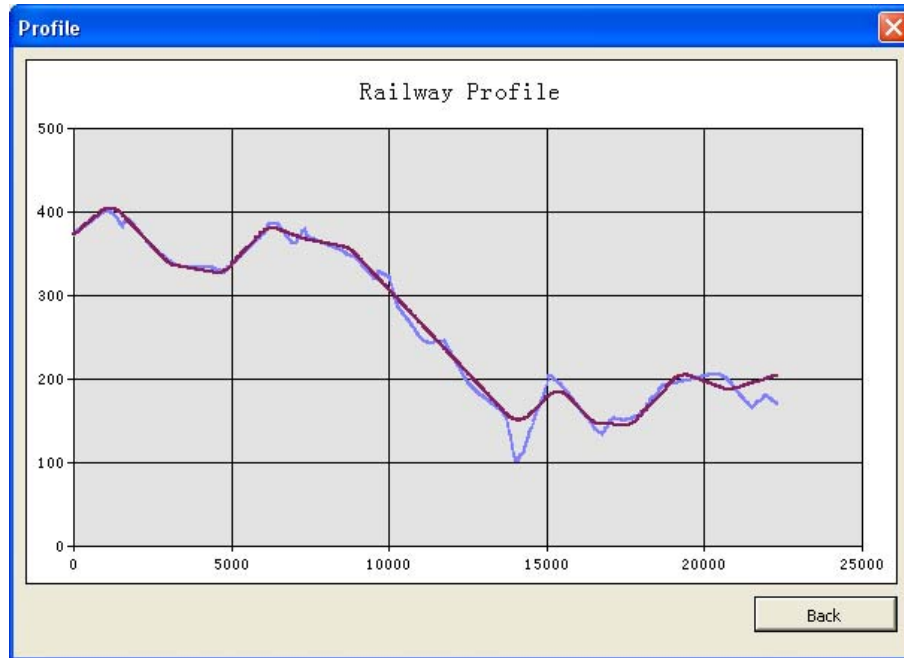
Figure 4.5 Alignment Profile Output

- Cross section output, which shows the cross section type (Cut-and-fill or bridge/tunnel) and the slopes from the edge of railway roadbed to the existing ground, as shown in Figure 4.6.
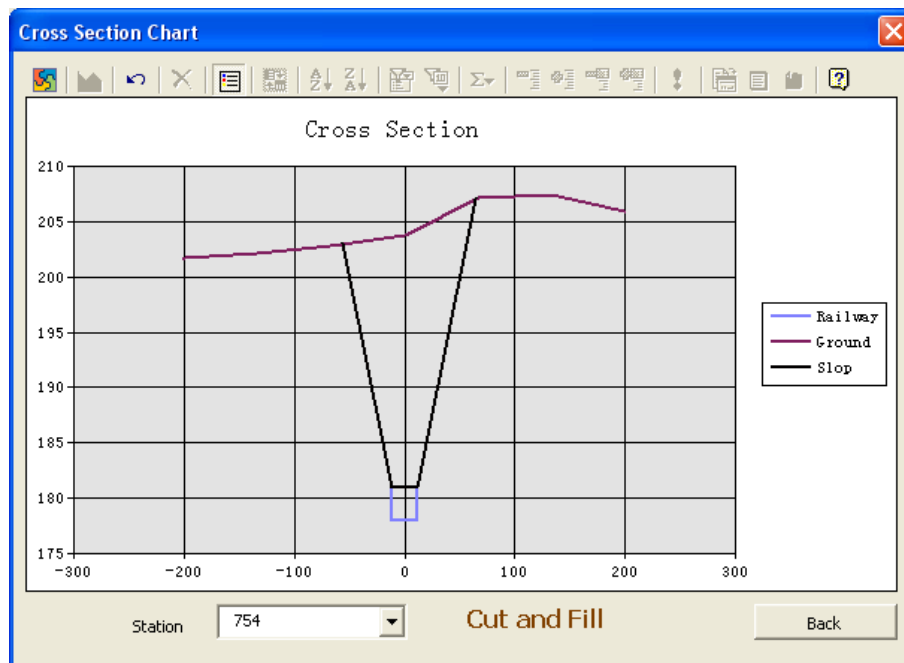
Figure 4.6 Cross Section Output

4.2 Case Study

4.2.1 Study Area

This section utilizes the real-world design scenario for the western section of the Baltimore Red Line as the study area in which we test the effectiveness of the proposed alignment optimization methodology.

As shown in Figure 4.7, this study area covers a 7-mile east-west transit corridor starting from the Social Security Administration. The corridor runs through I-70 Park-and-Ride, Edmondson Village Shopping Center and West Baltimore MARC Station, connecting the suburban residential area, shopping area and office parks. The alignment must end at the University of Maryland at Baltimore.

Figure 4.7 Illustration of Case Study Area

The local condition data of the study area were obtained from several agencies and incorporated into the background GIS database. These data include topological features, Maryland property distribution, county boundaries, and land-use patterns.

The proposed optimization program was also localized with the predefined design parameters as summarized in Table 4.1. These parameters are consistent with the guidelines from the Maryland Transit Administration [MTA, 2007].

Table 4.1 Design Parameters in the Case Study

Parameters	Description	Value	Unit
V	Design speed	45	mph
Δ_{BC}	Bridge clearance over water	10	ft
$N_{B,\min}$	Minimum number of consecutive bridge sections	5	each
$N_{T,\min}$	Minimum number of consecutive tunnel sections	5	each
Δ_B	Threshold of elevation difference at which a bridge becomes preferable to a fill section	40	ft
Δ_T	Threshold of elevation difference at which a tunnel becomes preferable to a cut section	40	ft
K_F	Unit cost for embankment	30	\$/cubic yard
K_C	Unit cost for excavation	30	\$/cubic yard
K_{TC}	Unit cost for tunnel excavation	200	\$/cubic yard
K_T	Track installation cost	300	\$/linear foot
$L_{T,\min}$	Minimum tangent length between curved sections	40	ft
$R_{C,\min}$	Minimum circular curve radius	500	ft
$l_{C,\min}$	Minimum circular curve length	45	ft
$E_{A,\max}$	Maximum applied superelevation	6	in
$l_{ST,\min}$	Minimum spiral length	40	ft
K_{runoff}	Factor for calculating minimum spiral length according to superelevation runoff rate	31	ft/in
K_{Jerk}	Factor for calculating minimum spiral length according to Jerk Rate	1.22	ft×h/(m×in)
K_{roll}	Factor for calculating minimum spiral length according to Vehicle roll rate	0.94	ft×h/(m×in)
$L_{\min}$	Minimum tangent length between vertical curves	50	ft
$g_{\max}$	Maximal grade	4	%
K_{VC}	Minimal vertical curve length per grade change	100	ft/%
$L_{VT,\min}$	Minimum vertical curve length	50	ft

4.2.2 Optimization Results

The case study was analyzed with the proposed computer program, using the embedded GIS database with local data and the predefined design parameters in Table 4.1. The population size was set at 100. The study also tested the following two different cutting plane settings of the optimization algorithm.

- Option 1: equally spaced cutting planes between two end terminals;

- Option 2: unequally spaced cutting planes between each pair of neighboring major transit stations.

The optimization results are organized into the following four parts:

4.2.2.1 Comparison of the Optimized Horizontal Alignments with the Empirical Alignment

Figure 4.8 compared the optimized alignments from both options to the empirically designed alignment, which is Alternative 4 from the Red Line Corridor Transit Study [MTA, 2008]. Alternative 4 is an LRT line and operates along Security Blvd to the I-70 Park-and-Ride and then along Cooks Lane to US 40. The alignment continues along US 40, turns to Martin Luther King Jr. Blvd, and reaches the University of Maryland at Baltimore. The empirical alignment mostly follows existing roadways.
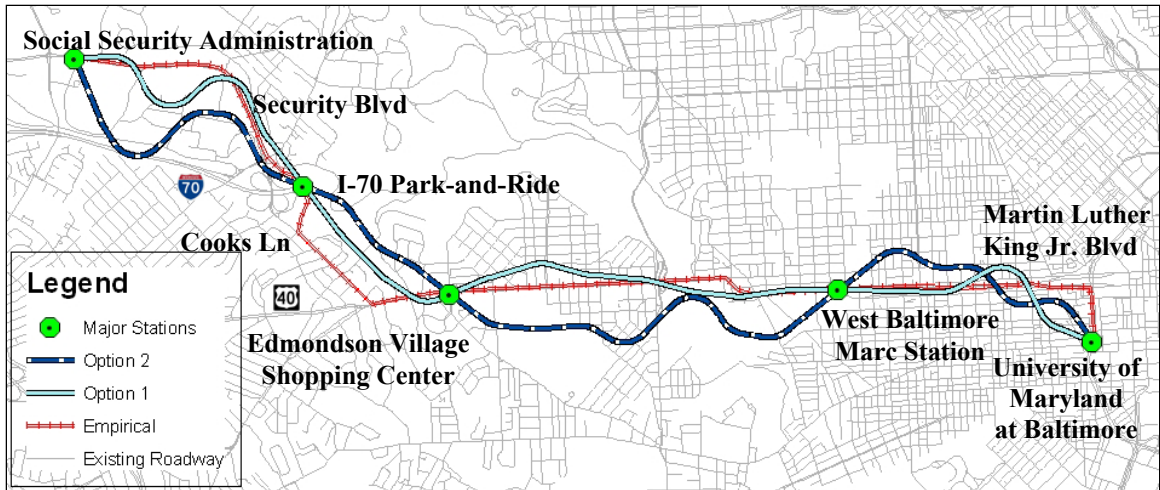
Figure 4.8 Horizontal Alignment Comparison

The proposed algorithm does not limit the alignment to the existing roadway network. Both options generate good alignments that minimize the total cost, satisfy all the geometry constraints, and precisely connect the major stations tangentially. The Option 1 alignment is more similar to the empirical one, using US 40 east of the Edmondson Village Shopping Center Station. On the west, the Option 1 alignment is straight through I-70 Park-and-Ride in order to shorten the track, instead of bypassing to Cooks Lane. Option 2 uses more curve sections to avoid some topological features, which decreases the bridge costs, but lengthens the track.

4.2.2.2 Comparison of the Optimized Profile with the Ground Elevation

This section compares the optimized profile from algorithm Option 1 with the ground elevation.

Figure 4.9 shows that the optimized profile tends to follow the ground elevation, while satisfying the grade and vertical curvature constraints. The algorithm

also accounts for the tradeoff between the total length of bridges and tunnels and other costs in order to decrease the total cost. Two bridges in the above optimized alignment have a total length of 1800 feet.
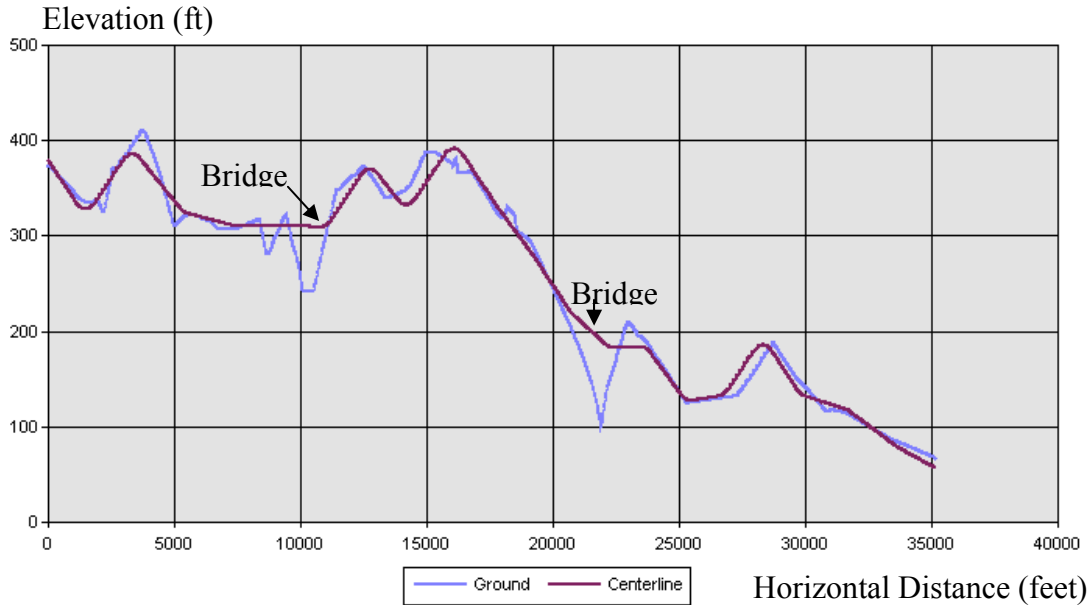
Figure 4.9 Optimized Profile vs. Ground Elevation

4.2.2.3 Comparison of the Cost Breakdown for the Optimized Alignments

Table 4.2 indicates that the two algorithm options generate alignments with relatively close total costs. The optimized alignment of Option 1 has slightly lower total cost, track total length, and cut/fill cost. The optimized alignment of Option 2 has lower bridge cost. The two alignments have almost the same ROW cost and no tunnel is used in either alignment.

Table 4.2 Cost Breakdown for Optimized Alignments (unit: million $)

	Earthwork Cost	Bridge Cost	Tunnel Cost	ROW Cost	Track Installation Cost	Total Cost
Option 1	19.6	6.0	0	11.7	10.5	47.8
Option 2	23.6	2.7	0	11.7	11.1	49.1

4.2.2.4 Comparison of the Efficiency of Algorithm Options

Table 4.3 compares the efficiency of the two algorithm options. These options require about the same time and number of generations to optimize the alignments. However, in the first generation that uses random values, Option 1 has more feasible alignments while the average cost exceeds that found with Option 2.

Table 4.3 Algorithm Efficiency

	Percentage of First Generation Feasible Alignments	First Generation Average Cost (million)	Number of Generations to the Solution
Option 1	57%	302	267
Option 2	14%	238	270

4.3 Sensitivity Analysis

Four categories of sensitivity analysis were conducted to examine the effectiveness of the proposed algorithm and the impact of different design parameters. As the two optimization options demonstrated similar performance in the previous section, only the first option was applied here to find the optimized alignment.

4.3.1 Impact of Introducing Bridge and Tunnel Calculation in the Total Cost

This section investigates the impact of incorporating bridge and tunnel calculation in the proposed algorithm by comparing the following two scenarios.

- SA-1: No bridge or tunnel is considered in the alignment. Cut-and-fill is applied to all alignment sections.

- SA-2: Use the method in section 3.2.2.1 to identify cut/fill sections and bridges/tunnel sections along the alignment. The following parameters are applied: $\Delta_B = 40$ feet, $\Delta_T = 40$ feet, $\Delta_{BC} = 10$ feet, $N_{B,min} = 5$, $N_{T,min} = 5$

Figure 4.10 compared the total cost and its breakdown for the optimized alignment in both scenarios.
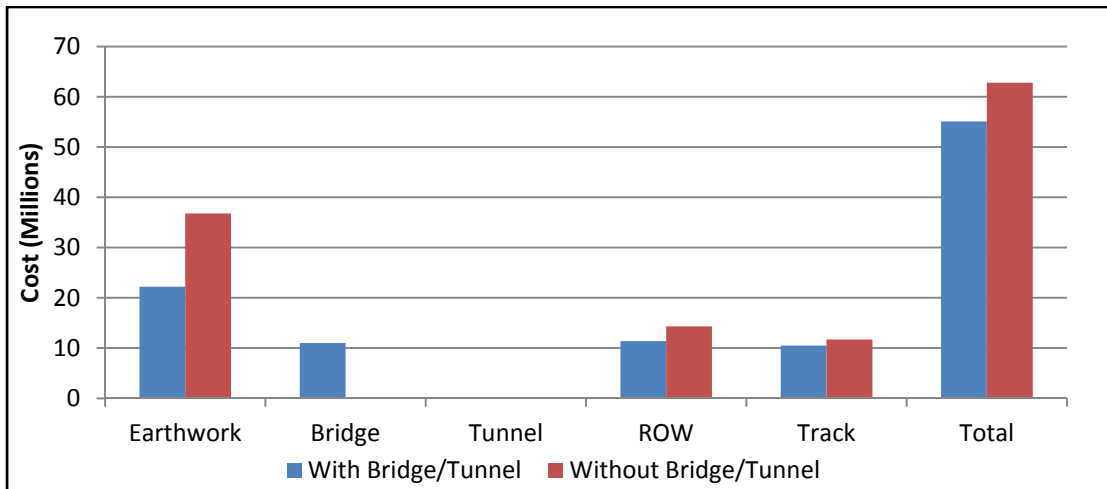
Figure 4.10 Cost with Bridge/Tunnel vs. Cost without Bridge/Tunnel

The above comparison shows that introducing bridge and tunnel options can significantly reduce the earthwork cost, and also reduce the ROW cost. Without using bridges and tunnels, the alignment tends to use more curve segments and the total track length increases, as shown in Figure 4.11. The total cost with bridge and tunnel options is 14% below the total cost without bridge/tunnel calculation.
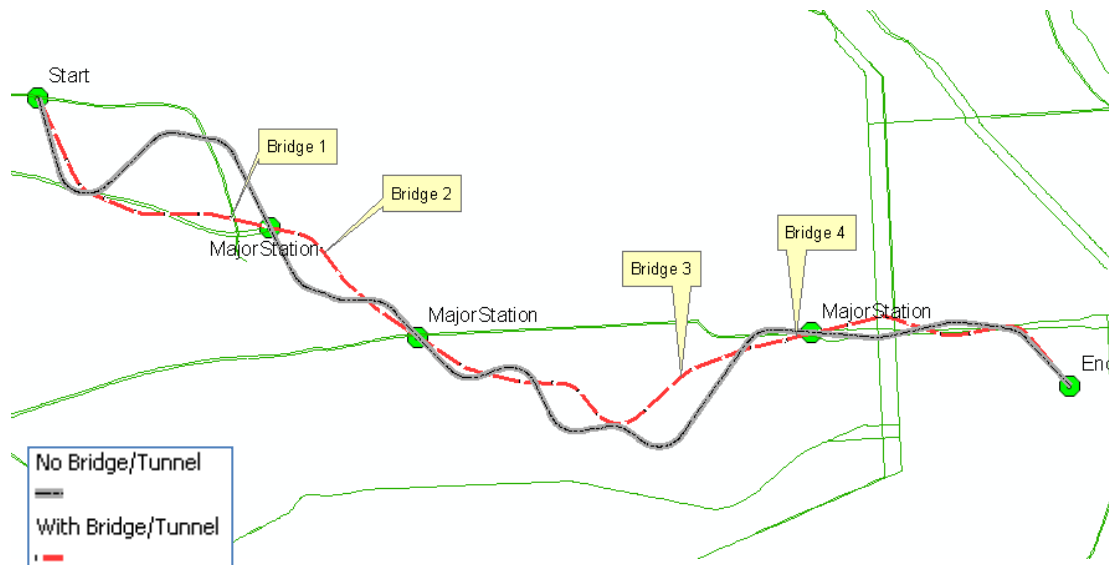
Figure 4.11 Horizontal Alignments with Bridge/Tunnel vs. without Bridge/Tunnel

4.3.2 Impact of Cutting Plane Spacing

Cutting planes are the vertical planes where PIs/VPIs of candidate rail transit alignments are located between two end terminals or neighboring stations. Cutting plane spacing determines the number of decision variables in the algorithm, in other words, the length of chromosome in GA. Longer spacing means fewer genes for an individual chromosome, which leads to a faster computation. Shorter spacing with more PIs, on the other hand, can provide more flexibility in the alignment, and thus a larger solution space. The sensitivity analysis on cutting plane spacing compares the following 6 parameter settings.

- SA 2-1: spacing = 1500 feet

- SA 2-2: spacing = 1600 feet

- SA 2-3: spacing =1700 feet

- SA 2-4: spacing =1800 feet

- SA 2-5: spacing =1900 feet

- SA 2-6: spacing =2000 feet

Figure 4.12 shows the optimized alignment cost with different cutting plane spacing and Figure 4.13 shows the percentage of feasible solutions in the first generation.
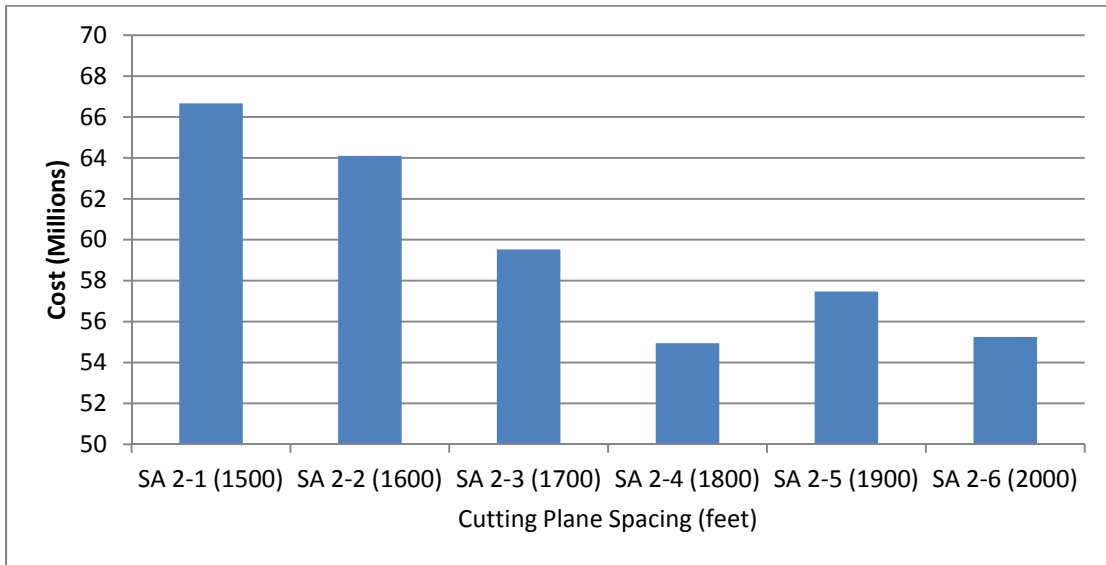
Figure 4.12 Optimized Costs with Different Cutting Plane Spacing
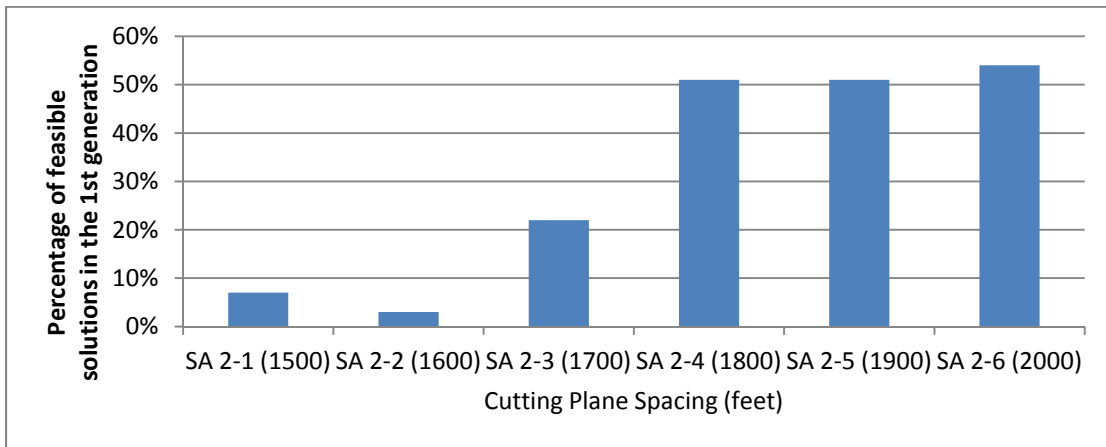
Figure 4.13 Percentage of 1st Generation Feasible Solutions with Different Spacing

The results lead to the following findings:

- When the spacing is in a good range (1700 to 2000 feet in the case study), the optimized alignment is not sensitive to the spacing. The lower spacing do not mean a lower cost. In the case study, the best alignment is found when the spacing is 1800 feet.

- When the spacing is too small, it is more difficult to find a feasible solution. It is easier for the randomly generated solutions to violate the geometry constraints on minimum tangent length, minimum curve radius/length, and minimum spiral length. In this study, less than 10% of the first generation individuals are feasible, with a cutting plane spacing of 1500 feet and 1600 feet.

- If the spacing is too large, there are too few PIs to efficiently and smoothly connect the major stations.

4.3.3 Impact of Railway Alignment Design Parameters

To test the algorithm sensitivity to the railway alignment design parameters, this section compared the optimized alignments with different design speed and different maximal design grade.

4.3.3.1 Design Speed

The sensitivity analysis on design speed includes the following 6 parameter settings:

- SA 3-1: speed = 60 mph

- SA 3-2: speed = 55 mph

- SA 3-3: speed = 50 mph

- SA 3-4: speed = 45 mph

- SA 3-5: speed = 40 mph

- SA 3-6: speed = 35 mph

Figure 4.14 shows the optimized alignment cost for each design speed and Figure 4.15 shows the percentage of feasible solutions in the first generation. These sensitivity analysis results indicate that it is more difficult to find feasible solutions when the design speed is higher. There is no feasible solution in the first 20 generations with 60 mph design speed. One reason for that is that the cutting plane spacing of 2000 feet is too close for high speed transit. It is also found that cost decreases as the design speed decreases. To connect the same major stations, the cost increases about by 34% when the design speed increases from 35 mph to 60 mph.
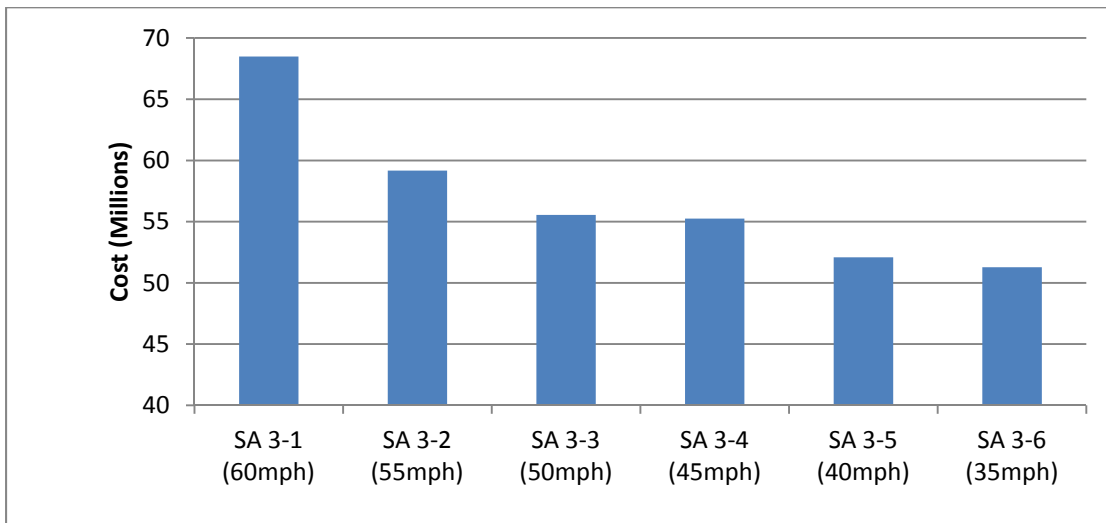
Figure 4.14 Optimized Costs with Different Design Speed
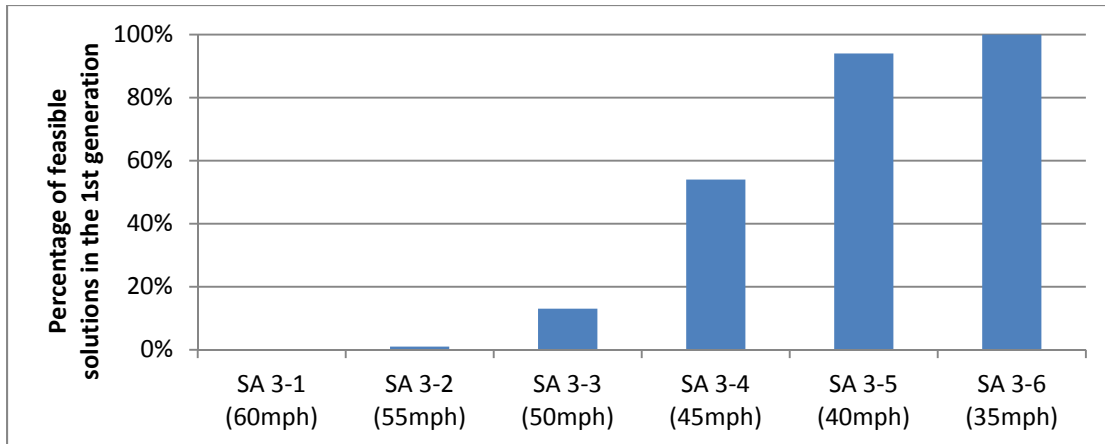
Figure 4.15 Percentage of 1st Generation Feasible Solutions with Different Design

Speed

4.3.3.2 Maximum Grade

The sensitivity analysis on maximal design grade includes the following 6 parameter settings:

- SA 4-1: grade = 2%

- SA 4-2: grade = 3%

- SA 4-3: grade = 4%

- SA 4-4: grade = 5%

- SA 4-5: grade = 6%

- SA 4-6: grade = 7%

Figure 4.16 shows the optimized total alignment cost and the cost of earthwork and bridge/tunnel. The results indicate that the total cost is much higher with a very strict maximum grade requirement. More earthwork and bridge/tunnel cost is needed to maintain a low grade. The total cost increases by 40% when the

maximum grade requirement changes from 4% to 2%. However, further relaxation of maximum grade constraint (grade>=4%) does not help reduce the construction cost in this case study.
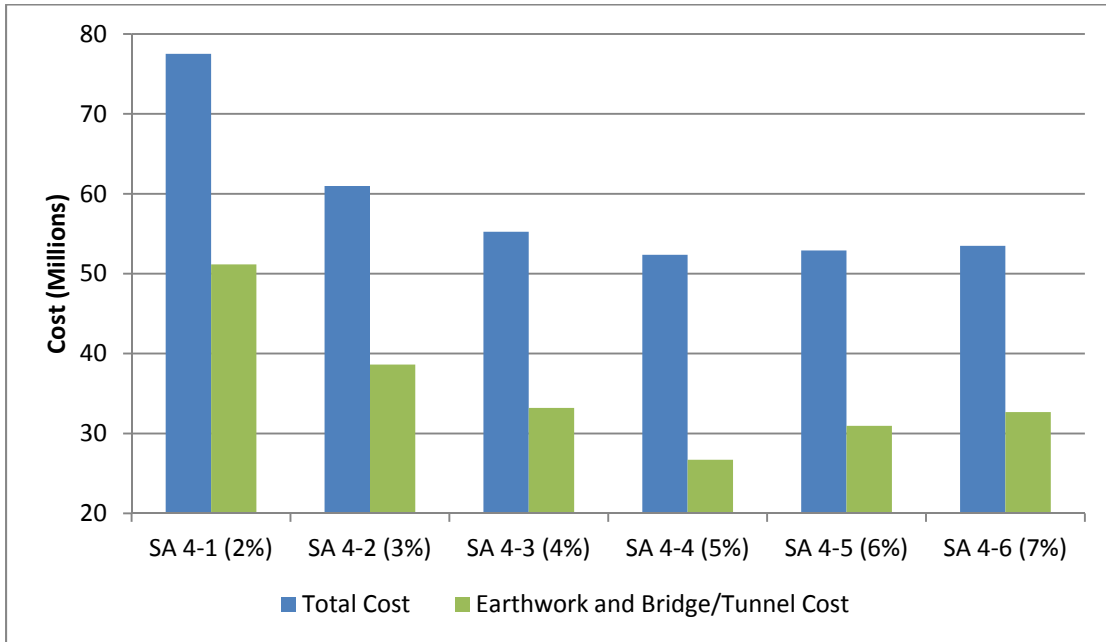
Figure 4.16 Optimized Costs with Different Maximum Grade

4.3.4 Impact of Genetic Algorithm Parameters

As with most machine learning algorithms, it is worth tuning the parameters such as population size, mutation rate, and crossover rate to find reasonable settings for the problem being worked on. This section compared the optimized alignments with different Genetic Algorithm parameters to test the algorithm's sensitivity to their values.

4.3.4.1 GA Population Size

The sensitivity analysis on population size includes the following four parameter settings:

- SA 5-1: population = 50

- SA 5-2: population = 100

- SA 5-3: population = 150

- SA 5-4: population = 200

The optimized alignment cost is shown in Figure 4.17. The algorithm can find a better solution within the same number of generations by using a larger population. The total cost decreases from 64 million to 49 million when population increases from 50 to 200. It is also found that further increases in population size beyond 150 tend to be less efficient to improve the solution.
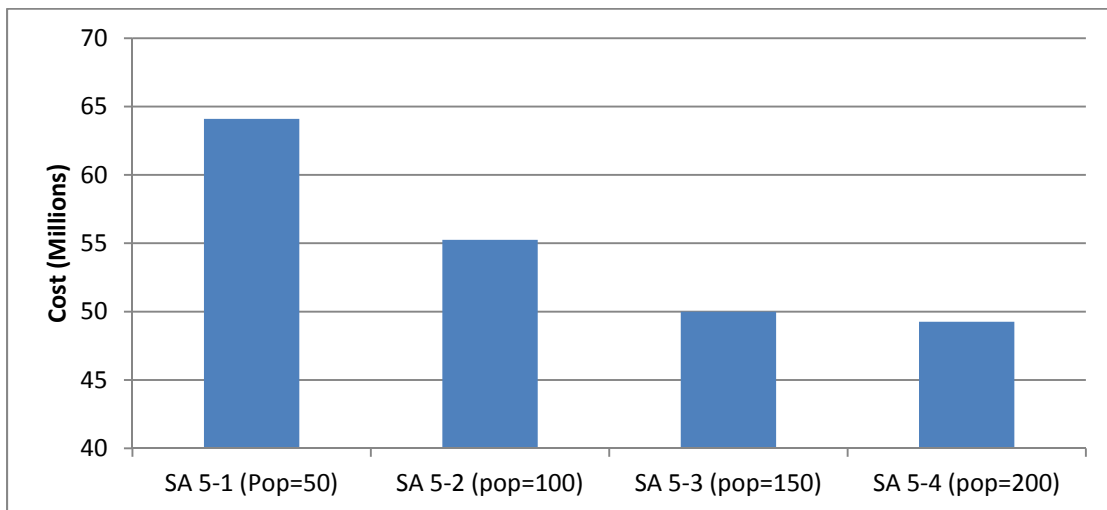
Figure 4.17 Optimized Costs with Different Population Size

4.3.4.2 GA Mutation Rate Sensitivity Analysis

The sensitivity analysis on mutation rate includes the following four parameter settings:

- SA 6-1: min/max mutation rate = 0.005/0.25

- SA 6-2: min/max mutation rate = 0.05/0.25

- SA 6-3: min/max mutation rate = 0.005/0.5

- SA 6-4: min/max mutation rate = 0.05/0.5

The optimized alignment cost is shown in Figure 4.18. The results indicate that an excessive mutation rate may lead to loss of good solutions. From the above figure, the best combination of mutation rates is 0.005/0.5.
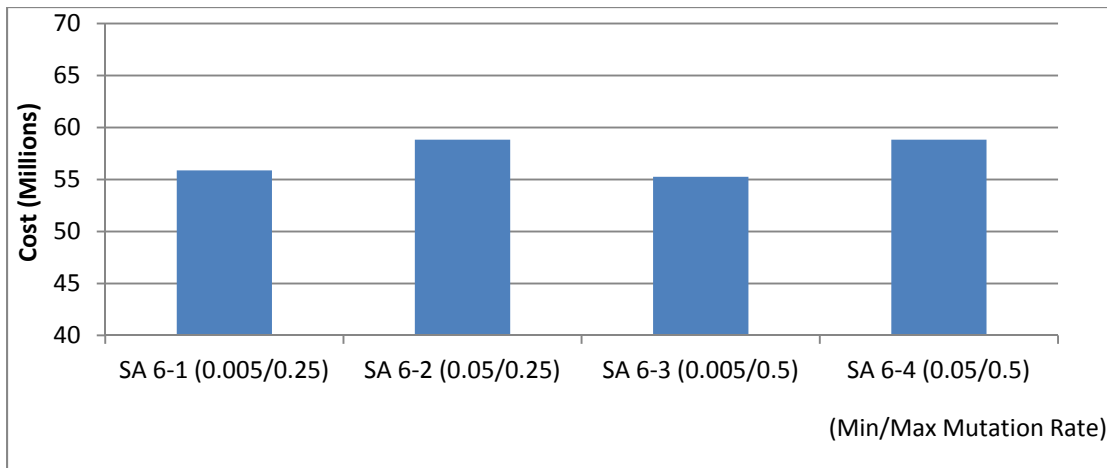
Figure 4.18 Optimized Costs with Different Mutation Rate Settings

4.3.4.3 GA Crossover Rate Sensitivity Analysis

The sensitivity analysis on crossover rate includes the following six parameter settings:

- SA 7-1: crossover rate = 0.5

- SA 7-2: crossover rate = 0.6

- SA 7-3: crossover rate = 0.7

- SA 7-4: crossover rate = 0.8

- SA 7-5: crossover rate = 0.9

- SA 7-6: crossover rate = 1.0

The optimized alignment cost is shown in Figure 4.19. The results show that the algorithm is not sensitive to the crossover rate when it is in the range of 0.5 to 0.9 for this case study. The best solution was found with a crossover rate of 0.6.
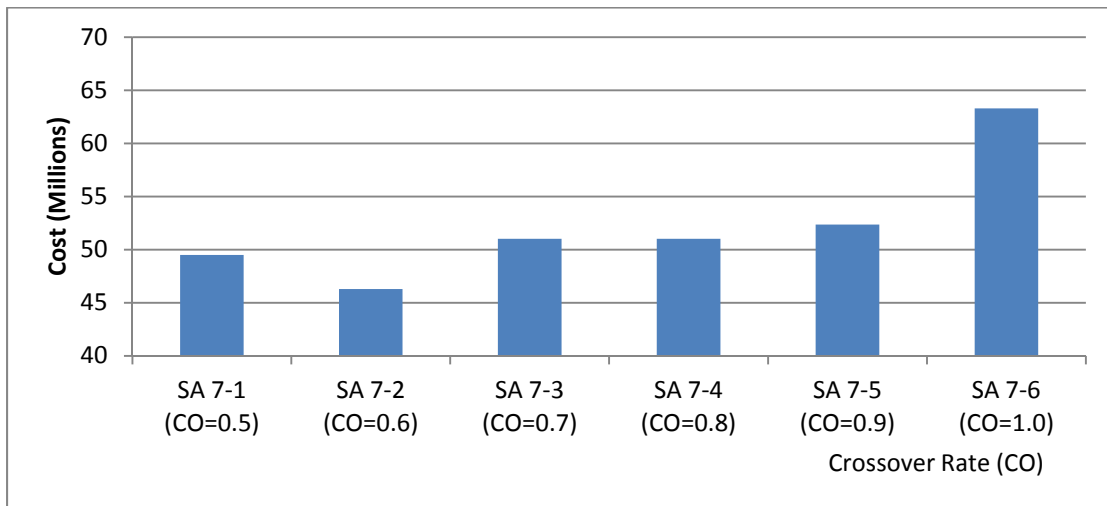
Figure 4.19 Optimized Costs with Different Crossover Rate

4.4 Variable Cutting Plane Spacing: Algorithm Improvement

The sensitivity analysis in Section 4.3 demonstrated that cutting plane spacing may greatly impact the efficiency and effectiveness of the solution algorithm. Thus, instead of requiring users to input the cutting plane spacing, this section presents an improvement to the algorithm which incorporates cutting plane spacing as decision variable in the optimization process. This also allows unequal spacing between neighboring cutting planes to introduce more flexibility in the candidate track alignments generated, while spacing between cutting planes in the base model has to be uniform in order to minimize user input.

4.4.1 Methodology

The methodology proposed here generates cutting planes with unequal spacing, while satisfying the minimal and maximal spacing constraints. These cutting plane locations are optimized along with the PI/VPI locations on them, using the GA based optimization heuristic.

4.4.1.1 Minimal and Maximal Spacing Constraints

Cutting plane spacing subjects to minimal and maximal spacing constraints. When the spacing is too short, it is more difficult to find a feasible solution, and easier to violate the geometry constraints on the alignment. When the spacing is too

long, the numbers of PIs between stations will be too few to connect the stations smoothly.

For each cutting plane spacing S_i (distance from cutting plane i-1 to cutting plane i), the constraint is

$$S_{min} \leq S_i \leq S_{max} \tag{4.2}$$

where S_{min} is the minimal spacing and S_{max} is the maximal spacing

The key factor considered in choosing minimal spacing S_{min} is the design speed. When the design speed is higher, longer spacing will be necessary to maintain the connectivity of horizontal alignments. For maximal spacing S_{max}, the algorithm here requires at least two PIs between any pair of neighboring stations.

4.4.1.2 Encoding of Chromosomes

The maximal number of cutting planes between the starting and ending terminals depends on the required minimal spacing and can be calculated as:

$$n_{max} = \lfloor d_{SE}/S_{min} \rfloor - 1 \tag{4.3}$$

where d_{SE} is the distance between the starting and ending terminals.

With cutting plane spacing as the additional decision variables, three variables are defined for each cutting plane. They are

- d_i : the horizontal distance of PI to the center of the cutting plane i

- Z_i : the vertical distance of PI to the center of the cutting plane i

- S_i : the spacing between the cutting plane i-1 and i

115

Including the elevations of all stations, the maximal length of a chromosome is equal to $(3 \times n_{max} + n_s) \times n_g$.

4.4.1.3 Decoding of Chromosomes to Cutting Plane Spacings

The chromosome is decoded into a series of normalized values between 0 and 1. The spacing $S_i = S_{min} + (S_{max} - S_{min}) \times \hat{S}_i$, where $\hat{S}_i$ is the normalized value.

The actual number of cutting planes n_{act} satisfies the following constraints, where the second constraint guarantees the maximal spacing constraint is met.

$$\sum_{i=1}^{n_{act}} S_i \leq d_{SE}$$

$$\sum_{i=1}^{n_{act}+1} S_i \geq d_{SE}$$

The following procedure is applied to revise $S_{n_{act}}$ to make sure the distance between the last cutting plane and the ending terminal also satisfies the minimal spacing constraint.

$$\text{if } d_{SE} - \sum_{i=1}^{n_{act}} S_i \leq S_{min} \text{ then}$$

$$S_{n_{act}}{}^{new} = S_{n_{act}} + S_{min} - (d_{SE} - \sum_{i=1}^{n_{act}} S_i)$$

After determining the cutting plane spacing, the rest of algorithm is unchanged from that in the base model.

To test the efficiency of the improved algorithm, this section applies the same design scenarios in section 4.3.2 that were used to analyze the impact of cutting plane spacing for the base algorithm.

4.4.2.1 Comparison of Optimized Costs

Figure 4.20 compares the optimized costs of the base algorithm with 6 uniform spacing settings (1500 to 2000 feet) with the optimized cost of the improved algorithm, which allows for variable spacings between 1500 feet to 3000 feet.
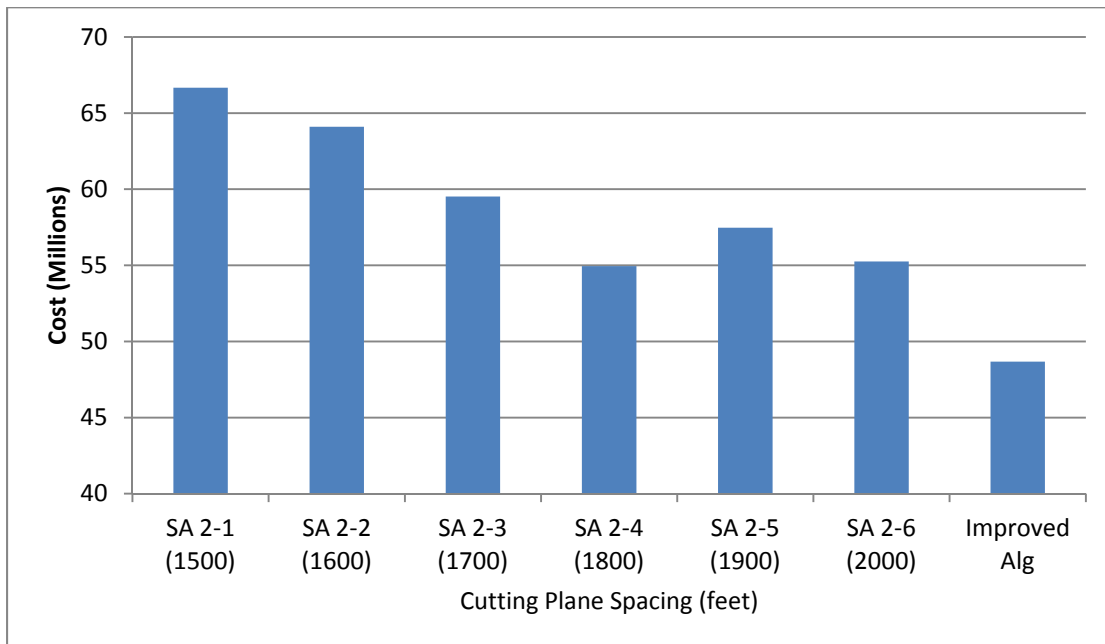
Figure 4.20 Comparison of Optimized Costs

The optimized cost from the improved algorithm is 48.7 million, 11.4% less than the lowest cost of 54.9 million from the base algorithm, which is obtained at cutting plane spacing of 1800 feet. This comparison demonstrates that, by allowing

more flexible cutting plane settings, the improved algorithm can search alignments in a larger solution space, and thus yield a better solution with lower costs.

4.4.2.2 Comparison of Convergence

Figure 4.21 compares the convergence of costs over successive GA iterations between the base algorithm and the improved algorithm.

Convergence: Base Algorithm vs. Improved Algorithm
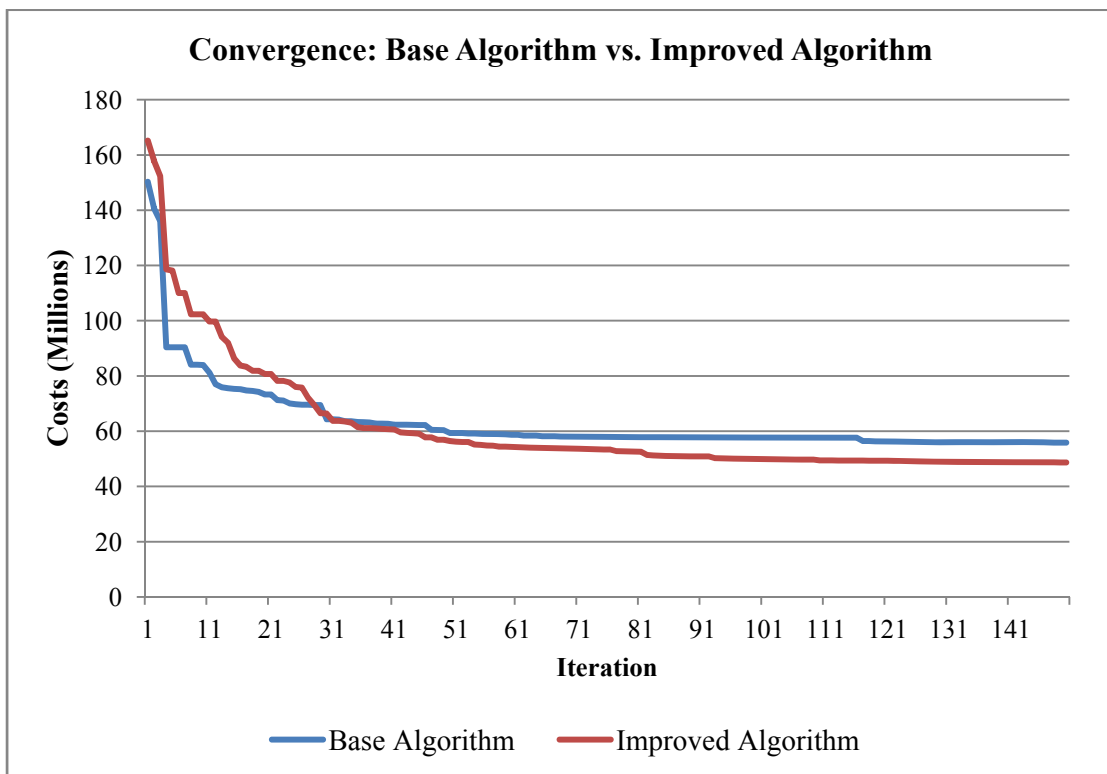
Figure 4.21 Comparison of Convergence

The comparison indicates that it is more difficult for the improved algorithm to find a feasible alignment with random cutting plane spacing in the initial iterations, so the costs are above those from the base algorithm. However, after some iterations

(about 30 iterations in this case study), the improved algorithm outperforms the base

algorithm.

4.4.2.3 Comparison of Optimized Horizontal Alignments

Figure 4.22 illustrates the optimized horizontal alignments generated by the

base algorithm and the improved algorithm.
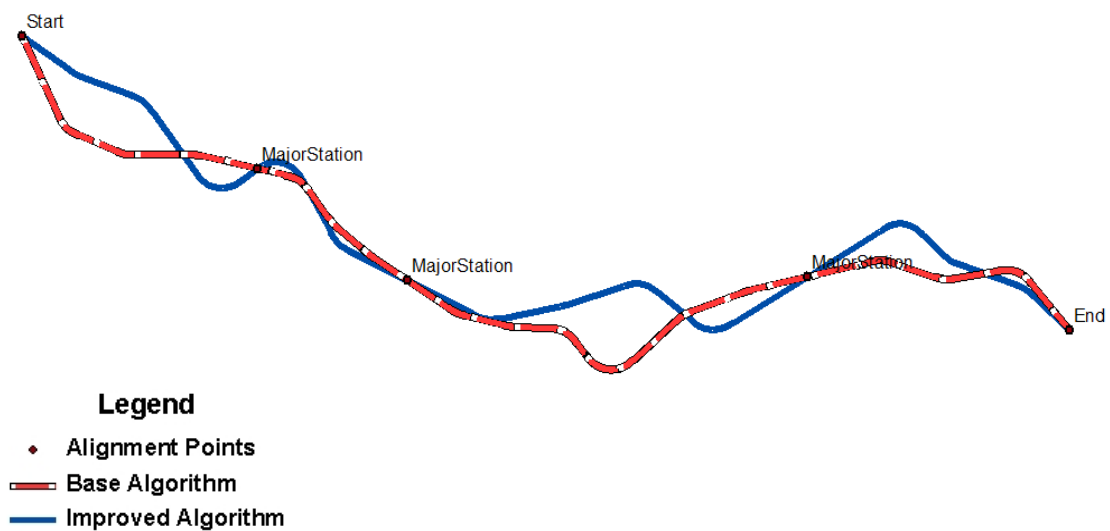
Figure 4.22 Comparison of the Horizontal Alignments

The comparison indicates that, in this case study, variable cutting planes of the

improved algorithm can lower the cost of optimized rail transit alignment by allowing

the alignment to utilize more curved segments to bypass high cost areas, such as those

with high right-of-way cost or earthwork cost.

4.5 Summary

This chapter presents a case study to demonstrate the effectiveness of the proposed rail transit alignment optimization methodology. This methodology aims to help engineers design track alignments that connect several major stations.

Section 4.1 describes a computer program that integrates the proposed optimization heuristic with a background GIS database. The program is programmed in Visual Basic and consists of five principal modules designed to collect user preferred design parameters, to search for the optimized rail transit alignment and to display the customized output of the optimized alignment. All modules are integrated by exchanging data inside the ArcGIS environment. The proposed module-based structure also offers the flexibility for further model updates or system expansion.

Section 4.2 presents an elaborate case study using the real-world design scenario for the western section of the Baltimore Red Line, which covers a 7-mile east-west transit corridor connecting five major stations. The proposed computer program is customized accordingly with various local data and MTA-required design parameters. The case study compares the optimized alignment profile with the ground elevation, the optimized horizontal alignments with the empirical alignment, and the cost breakdown and computation efficiency of two different cutting plane settings for the optimization algorithm. The numerical results demonstrates that, even in regions with complex topographical features, the proposed methodology can generate very good alignments that precisely connect the major stations tangentially, closely follow the ground elevation, and satisfy all the geometry constraints.

Extensive sensitivity analysis is conducted in section 4.3 to examine the impact of different design parameters and optimization parameters on the efficiency and effectiveness of the proposed methodology. The results indicate that the algorithm can achieve better performance by explicitly incorporating bridge and tunnel calculation, using proper cutting plane spacing, and tuning the Genetic Algorithm related optimization parameters to find reasonable settings. The optimized rail transit alignment is also affected by the critical design parameters, i.e., the design speed and maximal design grade.

Since the sensitivity analysis demonstrates that cutting plane spacing may greatly impact the efficiency and effectiveness of the solution algorithm, Section 4.4 presents an improvement to the algorithm which incorporates cutting plane spacing as decision variables in the optimization process. This allows unequal spacing between neighboring cutting planes and thus introduces more flexibility in the candidate track alignments generated.

Chapter 5: Optimizing Rail Transit Alignments

to Account for Vehicle Dynamics

In Chapter 3 a practical rail transit alignment optimization methodology is proposed, which can generate alignments that pass through preset station locations while meeting the special geometry constraints at these stations. The applicability of the methodology is extensively examined with a real-world case study and detailed sensitivity analysis in Chapter 4. This chapter presents an extension of the base model, which explicitly incorporates vehicle dynamics in the track alignment design. Such an extension aims to account for the significant impact of vehicle dynamics on operation and user cost in a rail transit system, and thus to generate alignments that better balance the initial cost with the operation and user costs recurring throughout the system's life cycle.

Section 5.1 presents the formulations of extended model, detailing a simulation process to realistically simulate the movement of trains along railway tracks and the need for dwell time, acceleration and deceleration at rail transit stations. The simulation yields more reliable estimates of travel time and energy consumption, which are two of the most critical parameters in calculating operation costs and user costs.

Section 5.2 presents a numerical study to demonstrate the essential trade-off among system costs, and its impacts on the design of rail transit alignments. A hypothetical topography scenario is created to illustrate the impact of vehicle

dynamics on the trade-offs among different system costs. The Baltimore Red Line is used as a case study to demonstrate that the model can find very good solutions in regions with complex topographies.

5.1 Model Formulations

This section presents the proposed rail transit alignment optimization methodology with vehicle dynamics. Kim and Schonfeld [1997] first investigated the impacts of vehicle dynamics on track alignment design while examining the benefits of dipped vertical alignments, which start getting lower upon leaving a station and gradually pick up elevation before the next station. Such dipped profiles between rail transit stations will take advantage of gravity for accelerating as well as decelerating trains, and thus reduce the operation cost by reducing break wear, saving energy, and decreasing travel time. However, a dipped profile may require additional earthwork and thus increase the construction cost. By incorporating vehicle dynamics into model formulations, the extended model in this chapter explicitly accounts for such tradeoff between operation cost and construction cost, and uses a GA to search for the optimized alignment.

The framework of the proposed rail transit alignment optimization methodology with vehicle dynamics is shown in Figure 5.1. The remaining sections will detail the two key steps, i.e., simulation of vehicle dynamics, and calculation of the cost function.
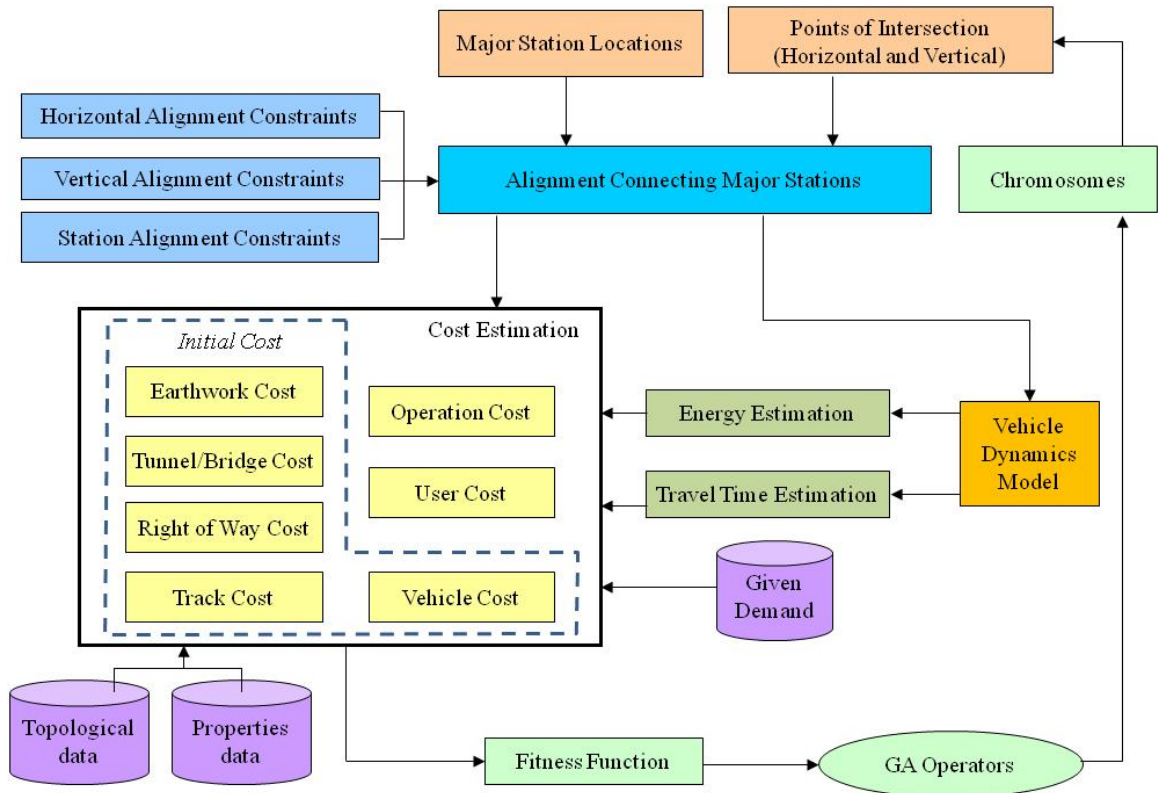
Figure 5.1 Framework of Rail Transit Alignment Optimization with Vehicle Dynamics

5.1.1 Simulation of Vehicle Dynamics

The vehicle dynamics model developed here is designed to analyze train energy consumption and travel time along a given track alignment. Between each pair of neighboring stations, a train typically experiences three distinctive stages of movement: acceleration from the previous station, cruising between stations, and braking to the next station. Sections 5.1.1.1 through 5.1.1.3 model train dynamics for these three stages as three iterative processes respectively, taking into account various factors including tractive/braking effort, resistance, rate of acceleration/deceleration, speed, energy consumption, and travel time. The flowchart for each of these iterative processes between a pair of neighboring stations is presented in the following Figure 5.2. The formulations are based on the essential train dynamics equations in Hay [1982]. Based on these iterative processes, Sections 5.2.1.4 presents the formulas to estimate the total round trip travel time and energy consumption.
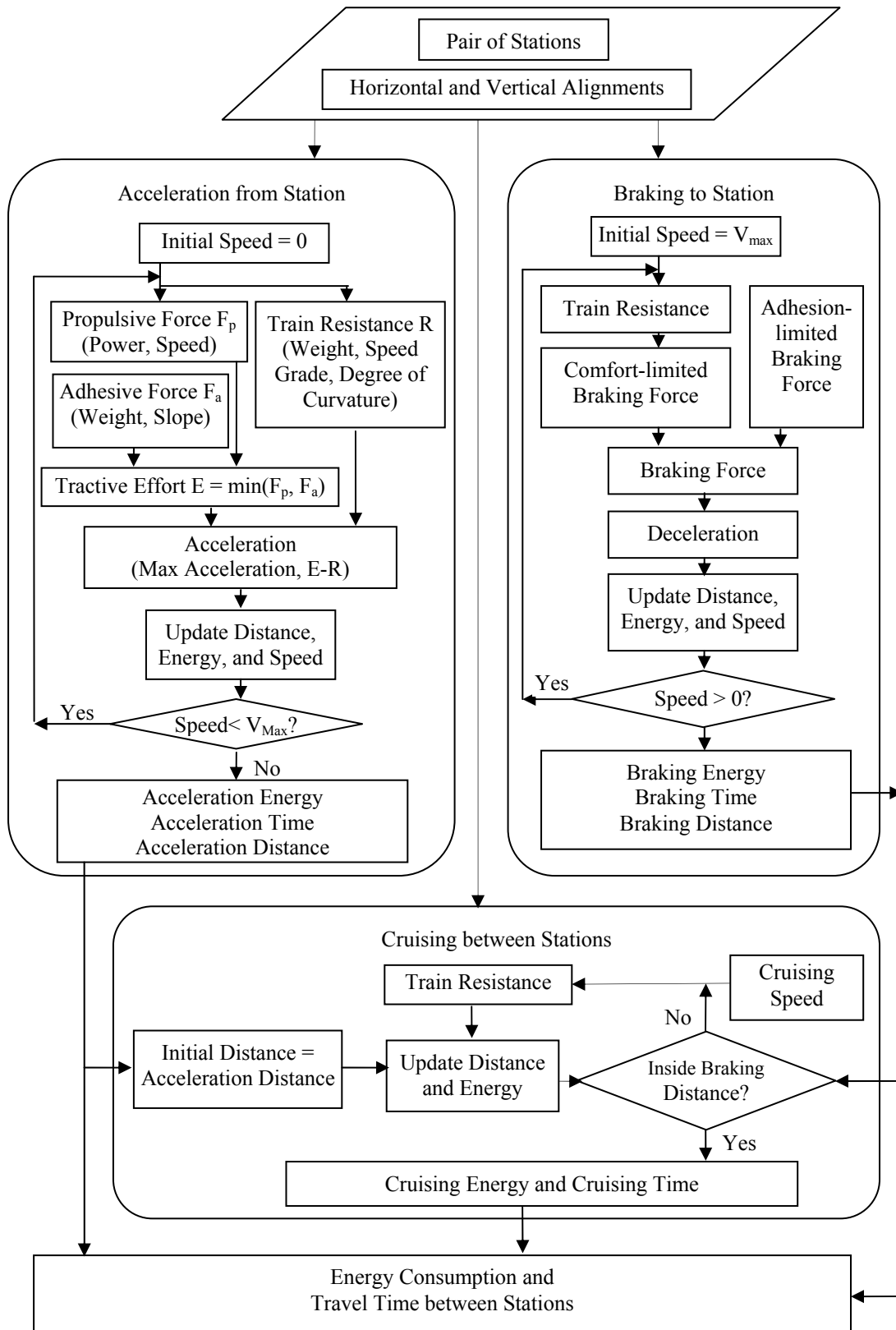
Figure 5.2 Three Stages Vehicle Dynamics

5.1.1.1 Acceleration Stage

The maximum tractive effort E_i is limited by available propulsive force F_{pi} at high speeds, and by adhesive force F_{ai} at low speeds. It is the minimum of these two forces:

$$E_i = \min(F_{pi}, F_{ai}) \tag{5.1}$$

Assuming the car is self-propelled, such as in an electric multiple unit (EMU), the propulsive force is:

$$F_{pi} = \frac{375\, \eta\, PN_c}{V_{i-1} \times 3600/5280} \tag{5.2}$$

where F_{pi}: propulsive force (lb) in interval i

η : transmission efficiency coefficient

P: power used to propel a car (hp)

N_c:number of cars per train

V_{i-1}: train speed (fps) in interval i-1

The adhesive force is:

$$F_{ai} = \mu_{i-1} WN_c \cos\theta_{i-1} \tag{5.3}$$

where μ_{i-1}: coefficient of friction

W: car weight (lb)

θ_{i-1}: angle of slope (radius) in interval i-1

Assuming a linear change between speed 0 kph and 80 kph, as in Figure 3.22 in [Vuchic, 1981], the friction coefficient can be obtained as:

$$\mu_{i-1} = 0.3 - \left[\frac{0.3-0.18}{80}\right] V_{i-1}(1.609) \tag{5.4}$$

The angle of slope is directly related to the gradient

$$\theta_{i-1} = \text{Arctan} \left(\frac{G_{i-1}}{100} \right) \tag{5.5}$$

where G_{i-1}: gradient in interval $i-1$, obtained from the vertical alignment.

The unit train resistance at time i in unit of lb/ton is formulated by employing the Davis Equation [Hay, 1982] as

$$R_{ui} = 1.3 + \frac{29n}{W/2000} + bV_{i-1} + \frac{CAV_{i-1}^2}{W/2000} + 20G_{i-1} + 0.8D_{i-1} \tag{5.6}$$

where R_{ui}: unit resistance of vehicle (lb/t) in interval i

n: number of axles per car

b: flange friction coefficient

C: air drag coefficient

A: cross-sectional area of train vehicle (ft^2)

D_{i-1}: degree of horizontal curvature in interval i-1, obtained from the horizontal alignment

In the above equation, the first component is bearing resistance depending purely upon the weight of the train. The second component is rolling resistance depending on the speed. The third component is aerodynamic resistance affected by speed, weight and train shape. The last two components are related to vertical gradient and horizontal curvature respectively.

The train resistance R_{vi} for each rail car is equal to the product of unit resistance and car weight in ton

$$R_{vi} = \frac{WR_{ui}}{2000} \tag{5.7}$$

The total resistance of a train R_{Ti} is the sum of the resistances of all cars in the train

$$R_{Ti} = \sum_{all\ v} R_{vi} \tag{5.8}$$

The acceleration force to move a train is equal to the tractive effort minus the resistance, which is also the product of acceleration rate and the train mass. The acceleration rate also needs to satisfy the max acceleration, which is based on both safety and passenger comfort concerns. The acceleration stage is complete when the train reaches its maximum design speed. Thus

$$a_i = \min \left(a_{max}, \frac{(E_i - R_{Ti})g}{\rho W N_c}, (V_{max} - V_{i-1})/\Delta t \right) \tag{5.9}$$

where a_i: acceleration in interval i (ft/s^2)

a_{max}: maximum allowable acceleration (ft/s^2)

g: gravity constant (ft/s^2)

ρ: coefficient for rotating masses

V_{max}: maximum speed constraint (fps)

Δt: simulation interval increments (s)

The speed and distance traveled in time interval i are

$$V_i = V_{i-1} + a_i \Delta t \tag{5.10}$$

$$\Delta d_i = V_{i-1} \Delta t + 0.5 a_i \Delta t^2 \tag{5.11}$$

The propulsive force energy consumption e_i in kwh in time interval i is

$$e_i = \left(\frac{a_i \rho W N_c}{g} + R_{Ti} \right) \Delta d_i / 2656000 \tag{5.12}$$

5.1.1.2 Braking Stage:

The actual braking force F_{bi}, which is used to decelerate the train, is the minimum of two forces

$$F_{bi} = \min (F_{bci}, F_{bai}) \tag{5.13}$$

where F_{bci} is comfort-limited braking force (lb) in interval i, which avoids exceeding deceleration limits for passenger comfort level; whereas F_{bai} is an adhesion-limited braking force

$$F_{bci} = \frac{a_{max}\rho W N_c}{g} - R_{Ti} \tag{5.14}$$

$$F_{bai} = \mu W N_c \cos\theta_{i-1} \tag{5.15}$$

The deceleration force to stop a train is equal to the braking force plus the resistance. After the braking stage, the train stops at the target station. The deceleration rate in time interval i is

$$a_i = \min \left(\frac{(F_{bi}+R_{Ti})g}{\rho W N_c}, \frac{V_{i-1}}{\Delta t}\right) \tag{5.16}$$

The speed and distance traveled in time interval i during braking stage are

$$V_i = V_{i-1} - a_i\Delta t \tag{5.17}$$

$$\Delta d_i = V_{i-1}\Delta t - 0.5a_i\Delta t^2 \tag{5.18}$$

The braking energy consumption in time interval i is

$$e_i = \left(\frac{a_i\rho W N_c}{g} - R_{Ti}\right)\Delta d_i/2656000 \tag{5.19}$$

5.1.1.3 Cruising Stage:

In the previous two stages, the acceleration distance from the previous station and the deceleration distance to the next station are calculated. Between these two

stages, the train seeks to maintain its cruising speed by applying the tractive force or braking force to balance the resistance. The speed, distance traveled, and the energy consumption in time interval i during this stage are

$$V_i = V_{max} \tag{5.20}$$

$$\Delta d_i = V_i \Delta t \tag{5.21}$$

$$e_i = |R_{Ti}| \Delta d_i / 2656000 \tag{5.22}$$

Equations 5.1 through 5.22 assume that tractive effort is distributed along the train and that all axles contribute equally to tractive effort and braking. It is also assumed that the control system, whether human or automatic, applies the specified speeds without error.

5.1.1.4 Travel Time and Energy Consumption for a Round Trip

Applying the above three-stage vehicle dynamics formulas to each segment along the train alignment, the total round trip travel time T_R in second can be calculated as follows, while including both running time and dwell time at terminals and stations.

$$T_R = \sum_{j \in acc} T_j + \sum_{j \in braking} T_j + \sum_{j \in cruising} T_j + 2D_e + 2D_m \tag{5.23}$$

where T_j:travel time on acceleration, braking, and cruising segments (s)

D_e:dwell time at end terminals (s)

D_m:dwell time at intermediate stations (s)

The total energy consumption E_R in kwh can be calculated as follows:

$$E_R = \sum_{j \in acc} \sum_i e_i + \sum_{j \in braking} \sum_i e_i + \sum_{j \in cruising} \sum_i e_i \tag{5.24}$$

5.1.2 Estimation of System Cost

The total cost C_{Total} used as the fitness function in the proposed heuristic is the sum of initial costs C_c, operation costs C_O, and user costs C_U.

$$C_{Total} = C_C + C_O + C_U \tag{5.25}$$

5.1.2.1 Initial Cost

The base model in Chapter 3 only considers construction cost C_c of the alignment in the optimization objective, which are capital costs and include earthwork costs C_E, bridges costs C_B, tunnels costs C_T, right-of-way costs C_R, track costs C_L, and train vehicle costs C_V:

$$C_c = C_E + C_B + C_T + C_R + C_L + C_V \tag{5.26}$$

All construction costs besides train vehicle costs have been discussed in the previous chapter. Assuming a fixed headway H in train schedule, the number of trains N_T can be calculated as the round trip travel time divided by the headway:

$$N_T = \left\lceil \frac{T_R}{H} \right\rceil \tag{5.27}$$

The vehicle costs C_V are the product of the number of trains N_T, number of cars per train N_c, and the cost for a train car K_V in millions:

$$C_V = N_T \times N_c \times K_V \tag{5.28}$$

The above formulations of train vehicle costs indicate that shorter travel time leads to fewer required trains, and thus decreases the cost of purchasing vehicles.

5.1.2.2 Operation Cost and User Cost

Unlike the aforementioned capital costs that occur only at the time of purchase or construction, the operation and user costs of a rail transit system occur throughout the life of the urban rail transit system.

The operation costs are modeled here to include two components, energy costs and other operation and maintenance costs. Daily energy costs are the product of the number of round trips, round trip energy consumption E_R obtained from the previous section, and the unit cost of energy K_e in \$/kwh. Other operation and maintenance costs are assumed as linear to the total passenger miles the railway system carries based on user projected travel demands. Thus the annual operation costs A_o in millions are:

$$A_o = 365 \left\lceil \frac{T_O}{H} \right\rceil \times E_R \times \frac{K_e}{10^6} + 365 \sum_i \sum_j D_{i,j} L_{i,j} \times \frac{K_O}{10^6} \qquad (5.29)$$

where T_O:train operating time per day (s)

$D_{i,j}$:daily demand from station i to station j

$L_{i,j}$:distance from station i to station j (miles)

K_O:unit operation and maintenance cost (\$/passenger-mile)

Assuming a fixed annual interest rate r and number of years n_a for the life cycle analysis, the present value of operating costs is

$$C_o = A_o \frac{(1+r)^{na}-1}{r(1+r)^{na}} \qquad (5.30)$$

The user costs are the costs rail transit system users pay for utilizing rail transit service, which include the travel time costs, transit fare, and access costs.

Since demands and station locations are given, transit fare and access costs are fixed too. Thus this model only considers travel time costs in the optimization process.. The annual user costs A_U in millions are

$$A_U = 365 \sum_i \sum_j D_{i,j} \times T_{i,j} \times \frac{1}{3600} \times \frac{K_U}{10^6} \qquad (5.31)$$

where $T_{i,j}$:travel time from station i to station j (s)

K_U:unit user cost ($/passenger-hour)

Applying the same life cycle analysis method as for operation costs, the present value of user costs C_U in the rail transit life cycle is

$$C_U = A_U \frac{(1+r)^{na}-1}{r(1+r)^{na}} \qquad (5.32)$$

5.2 Case Study

The numerical study in this section includes two scenarios:

- A hypothetical topography scenario is created to illustrate the impact of vehicle dynamics on the trade-off among different system costs.

- The Baltimore Red Line is used to demonstrate the model's applicability in real-world practice which involves complex topographical features.

Three optimized alignments are generated by setting different objectives as the fitness function in the proposed method.

- Alignment A1: minimize only the initial costs;

- Alignment A2: minimize only the operation and user costs;

- Alignment A3: minimize the total costs.

This thesis excludes ROW costs in the case study in order to eliminate the impacts of random location-specific costs and allow consistent comparisons of the continuous alignment-related costs.

5.2.1 Hypothetical Topography Scenario

5.2.1.1 Scenario Design

The purpose of this hypothetical topography scenario is to demonstrate the model's ability to jointly minimize travel time and energy consumption with the track alignment optimization, and to demonstrate the impact of vehicle dynamics on the trade-off among initial costs, operation costs, and user costs.

This scenario assumes a completely flat terrain in the study area, where all surface features have a fixed elevation of 500 feet above sea level. The earthwork cost (for example to raise stations above the terrain) is computed within the total cost. The two terminals are 5 miles apart with two predetermined intermediate stations. A 4-car passenger train is considered with a maximum acceleration rate of 4.265ft/s2, and a power of 697 hp/car. Train maximum speed is 75 mph. The unit cost for cut/fill earthwork is $15 per cu ft. and the unit energy cost is $0.485/kwh. The unit operation and maintenance cost and unit user cost are $0.16/passenger-mile and $11.5/passenger-hour, respectively. Annual interest rate in this scenario is 2%.

5.2.1.2 Numerical Results

Figure 5.3 compares the vertical and speed profiles of the three optimized alignments.
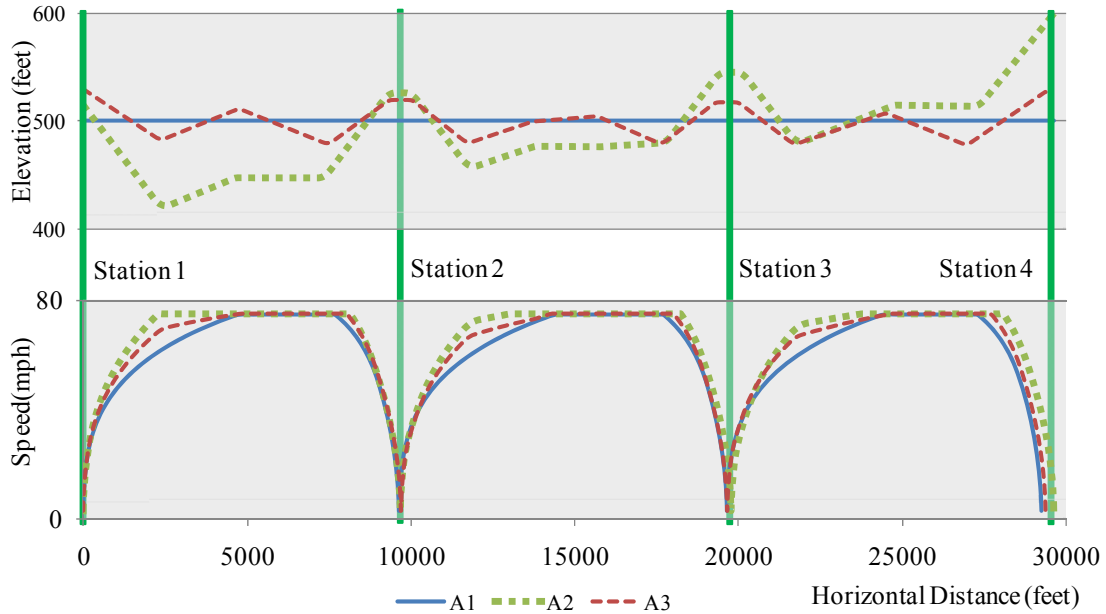
Figure 5.3 Comparison of three vertical and speed profiles.

The A1 profile, seeking to minimize the initial cost (mainly earthwork cost here), follows the flat ground surface. The A2 profile, in minimizing the operation and user cost, has the steepest dipped profile in order to reduce the energy consumption and travel time. The A3 profile, minimizing the total cost, is a compromise between alignments A1 and A2, and represents the trade-off between the energy and travel time savings and extra earthwork.

Table 5.1 compares the cost, travel time, and energy consumption of A1, A2 and A3.

Table 5.1 Comparison of Three Optimized Alignments

	Alignment A1	Alignment A2	Alignment A3
Earthwork Cost (Millions of $)	0.0	51.0	7.8
Operation Cost (Millions)	105.3	91.4	95.9
User Cost (Millions)	45.6	44.1	44.6
Operation + User Cost (Millions)	150.9	135.5	140.5
Total Cost (Millions)	169.3	205.0	166.7
Energy Consumption (kwh)	336	275	295
Travel Time (seconds)	744	712	723

When the fitness function only considers initial cost, the optimized alignment A1 has no earthwork cost because flat terrain is assumed. However, A1 also has the highest energy consumption and travel time, thus having the highest operation cost and user cost among the three alignments.

When the fitness function only considers operation cost and user cost, the optimized alignment A2 has the lowest operation and user cost of $135.5M. However, in reducing energy consumption by 18.2% and travel time by 4.3%, alignment A2 increases earthwork cost by $51M, which raises the total cost to $205M, the highest among the three alignments.

Alignment A3, in which the total cost is minimized, represents trade-offs between alignments A1 and A2. Compared to A1, the reductions of round trip travel time and energy consumption of alignment A3 are 2.8% and 12.2%, respectively, which are less than the reductions of A2. Because A3 is less focused than A2 on searching for minimal operation and user cost, the extra earthwork cost $7.8M in A3 is much less than $51M in A2. The initial cost, operation cost and user cost of A3 are all between those costs in A1 and A2. The total cost of A3 is $166.7M, the lowest

among the three alignments, which demonstrates a better balance between low energy consumption/travel time and low initial costs.

5.2.1.3 Sensitivity Analysis

Sensitivity analysis (SA) is performed here to explore how the model's optimization results depend on: (1) maximum train speed; (2) unit price of energy cost; and (3) interest rate. The following ranges of values are used in SA:

- Maximum train speed: 45 mph to 75 mph (every 10 mph, 4 increments)

- Unit price of energy cost: $0.1/kwh to $0.6/kwh (every $0.1/kwh, 6 increments)

- Interest rate: 2% to 8% (every 2%, 4 increments)

Table 5.2 shows the sensitivity analysis results for the effects of the above three parameters.

Table 5.2 Sensitivity Analysis Results

Parameter		Round Trip Energy Consumption (khw)			Round Trip Travel Time (seconds)			C_O+C_U (M)	C_E (M)	C_{Total} (M)
		A1	A3	(A3-A1)/A1	A1	A3	(A3-A1)/A1	A3-A1	A3-A1	A3-A1
Max Cruising Speed (mph)	45	144	142	-1.4%	988	986	-0.2%	-0.43	0.19	-0.24
	55	203	199	-2.0%	858	856	-0.2%	-1.16	0.64	-0.52
	65	271	256	-5.5%	782	774	-1.0%	-3.70	2.83	-0.87
	75	336	295	-12.2%	744	723	-2.8%	-10.45	7.78	-2.67
Unit Energy Cost ($/kwh)	0.1	336	336	0.0%	744	742	-0.3%	-0.06	0.01	-0.05
	0.2	336	329	-2.1%	744	738	-0.8%	-0.94	0.83	-0.09
	0.3	336	326	-3.0%	744	737	-0.9%	-1.71	1.32	-0.39
	0.4	336	305	-9.2%	744	727	-2.3%	-6.47	5.34	-1.13
	0.5	336	295	-12.2%	744	723	-2.8%	-10.45	7.78	-2.67
	0.6	336	291	-13.4%	744	722	-3.0%	-13.58	8.86	-4.72
Interest Rate	2%	336	295	-12.2%	744	723	-2.8%	-10.45	7.78	-2.67
	4%	336	305	-9.2%	744	728	-2.2%	-5.65	5.13	-0.52
	6%	336	323	-3.9%	744	737	-0.9%	-1.80	1.35	-0.45
	8%	336	330	-1.8%	744	738	-0.8%	-0.75	0.70	-0.05

When considering vehicle dynamics in the alignment optimization, as the maximum train speed increases from 45 mph to 75 mph, the energy consumption reduction varies from 1.4% to 12.2%, and travel time reduction also varies from 0.2% to 2.8%. The savings of operation costs and user costs increase from $0.43M to $10.45M, whereas the earthwork costs increases from $0.19M to $7.78M. After the trade-off, the total cost saving is $0.24M at the lower speed of 45 mph, and $2.67M at 75 mph.

As the energy cost increases from $0.1/khw to 0.6/khw, the energy savings achieved by considering vehicle dynamics in the alignment optimization increase from 0.0% to 13.4%, and travel time savings also rise from 0.3% to 3.0%. The savings of operation costs and user costs increase from $0.1M to $13.6M, while the earthwork costs also increases from 0 to $8.9M.

As the interest rate increases from 2% to 8%, the energy consumption reduction due to consideration of vehicle dynamics decreases from 12.2% to 1.8%, and the travel time reduction also decreases from 2.8% to 0.8%.

5.2.2 Baltimore Red Line

This section utilizes the design scenario from the case study of the base model to test the impacts of considering vehicle dynamics and additional cost components (i.e. operation costs and user costs) in this chapter's extended model.

Table 5.3 presents the demand matrix of the station-to-station personal trips among the five preset stations, which is assumed to be available from external travel

forecasts. Table 5.4 presents the values of various design parameters used in the case study.

Table 5.3 Demand Matrix

	Station 1	Station 2	Station 3	Station 4	Station 5
Station 1	0	100	200	400	800
Station 2	100	0	100	200	400
Station 3	200	100	0	100	200
Station 4	400	200	100	0	100
Station 5	800	400	200	100	0

Table 5.4 Design Parameters

Parameters	Description	Value	Unit
a_{max}	maximum allowable acceleration	± 4.265	ft/s^2
A	cross-sectional area of train vehicle	113	ft^2
C	air drag coefficient	0.0007	
D_e	dwell time at end terminals	300	s
D_m	dwell time at intermediate stations	60	s
g	gravity constant	32.2	ft/s^2
H	train headway	900	s
K_e	unit cost of electricity energy	0.097	\$/kwh
K_O	unit operation and maintenance cost	0.48	\$/passenger-mile
K_U	unit user cost	11.5	\$/passenger-hour
K_V	cost for a train car	0.1	million
n	number of axles per car	4	
n_a	number of years for the life cycle analysis	40	years
N_c	number of cars per train	4	
P	propulsive power per car	697	hp
r	annual interest rate	5%	
Δt	simulation interval increments	1	s
T_O	train operating time per day	12	hours
V_{max}	maximum speed	45	mph
W	car weight	40	ton
η	transmission efficiency coefficient	0.82	
ρ	coefficient for rotating masses	1.05	

5.2.2.1 Comparison of the Horizontal Alignments

Figure 5.4 illustrates the three alternative horizontal alignments generated from the proposed methodology with different design objectives.

When the model only seeks to minimize initial costs, more curved segments are used to avoid extreme high/low topological features and to reduce the earthwork, as shown in Alignment A1. Curves with short radius and reverse curves are also observed in this alternative alignment that minimizes only the initial costs.
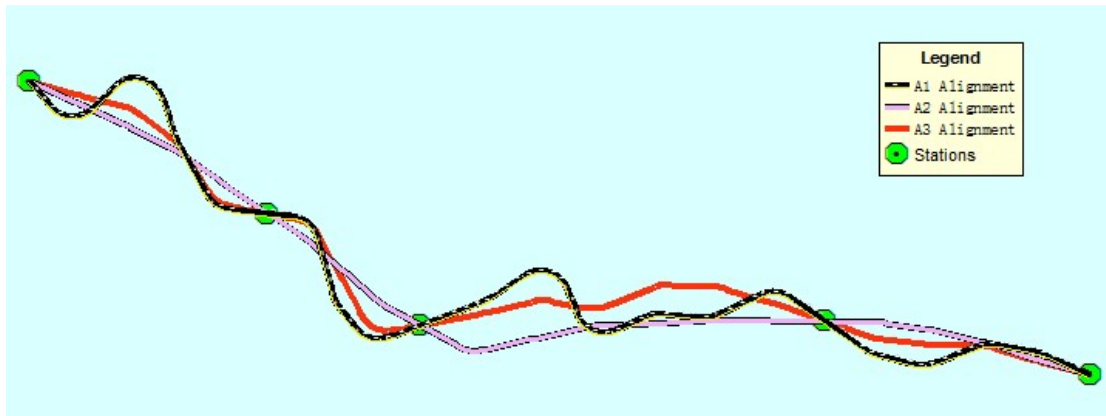
Figure 5.4 Comparison of Horizontal Alignments

When the model only seeks to minimize the operation and user costs, it seeks the shortest and fastest alignment to connect all stations while satisfying all geometric constraints, as shown in Alignment A2. This optimized alignment has very few curves and those curves used here to maintain the alignment's smoothness usually have a long radius. Travel time is the least when the objectives are to minimize operation and user costs.

The alternative Alignment A3 represents the trade-off between the two sets of objectives. The alignment is somewhere between the above two extreme cases. It has more curves than A2, and fewer than A1.

5.2.2.2 Comparison of the Vertical Profile and Vehicle Dynamics

Figure 5.5 illustrates the three alternative vertical alignments, or profiles, generated from the proposed methodology with different design objectives. The ground profiles for these three alternatives follow the horizontal alignments shown in Figure 5.4. Hence, the ground profiles differ for the three alignments.

The A1 profile uses grade changes more often to closely match the ground elevation changes and to minimize the earthwork and initial cost.

A2 has the minimal changes of grade, indicating a smoother vertical alignment to reduce the energy consumption.

A3 represents the trade-off between the grade changes and the extra earthwork for a smoother vertical alignment. The round trip time of Alignment A3 is 36.5 minutes, slightly above 35.7 minutes for A2, and slightly below 38.5 minutes for A1. The energy consumption of 225 kwh per round trip for A3 is also between 346 kwh for A1 and 225 kwh for A2. Thus, Alignment A3 demonstrates a better balance between energy consumption and initial costs.

Elevation (feet)

A1 Profile

Horizontal
Distance (feet)

Elevation (feet)

A2 Profile

Horizontal
Distance (feet)

Elevation (feet)

A3 Profile

Horizontal
Distance (feet)
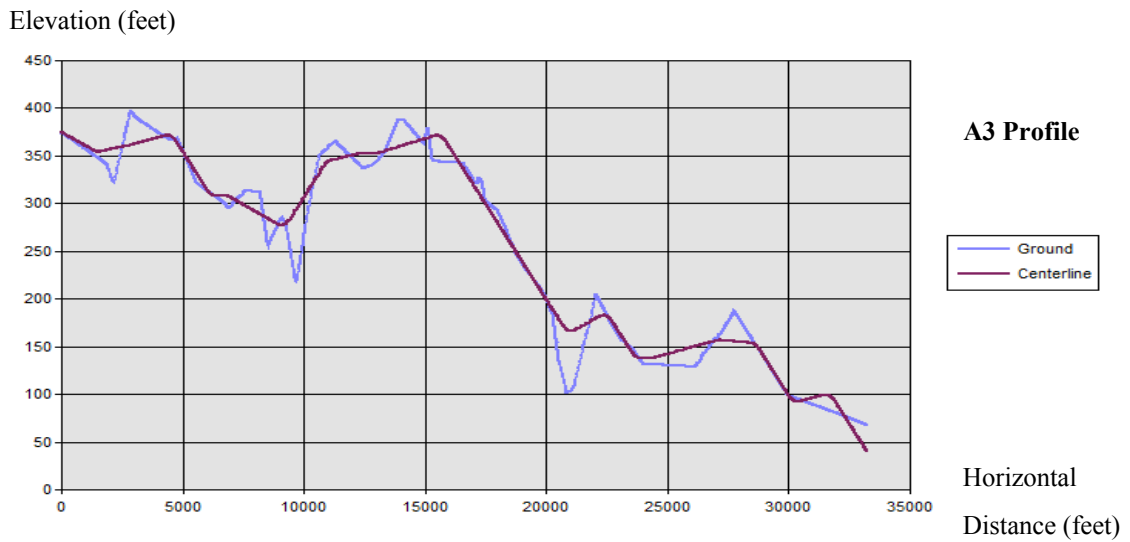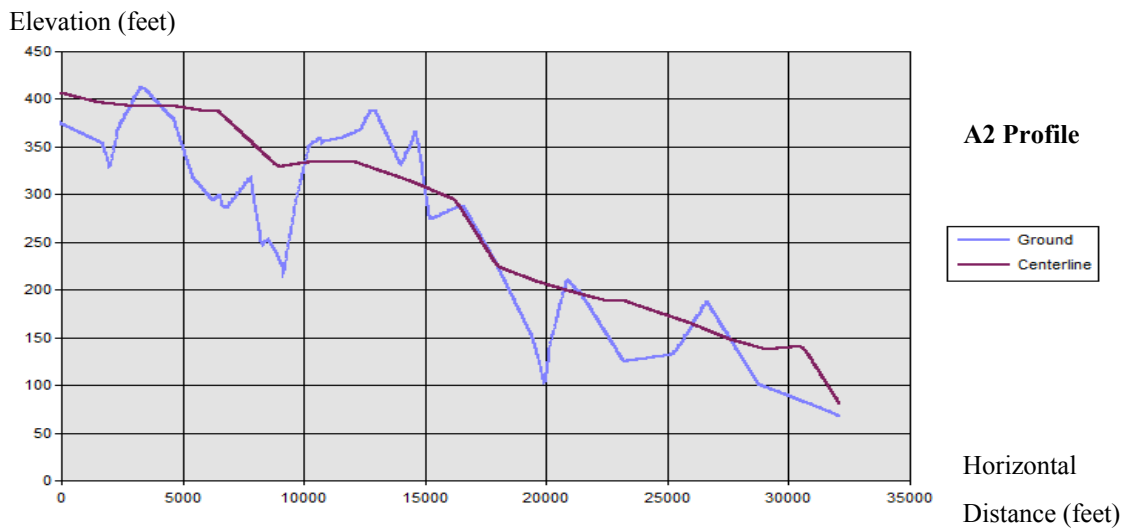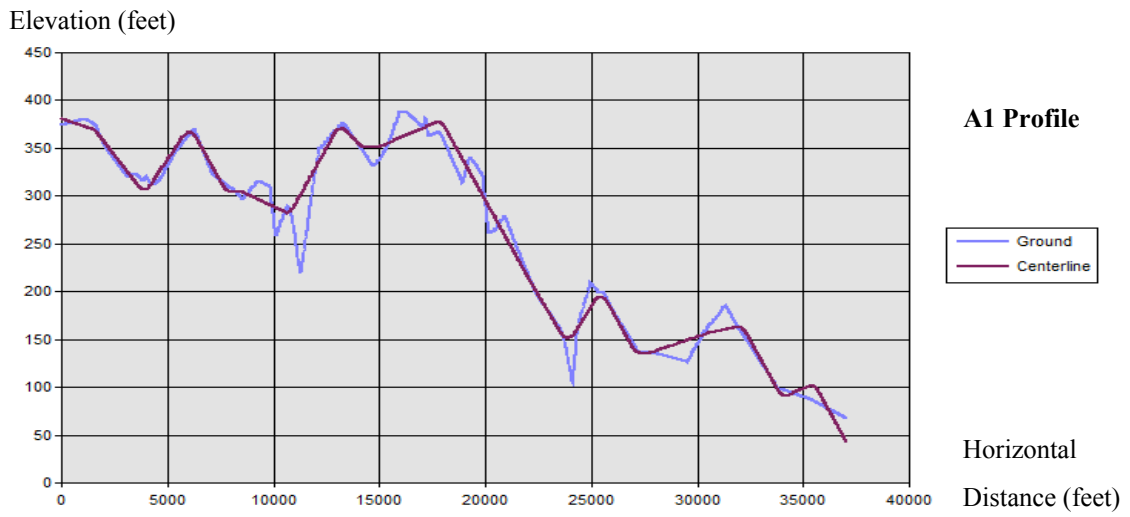
Figure 5.5 Comparison of Profiles

5.2.2.3 Comparison of Cost Breakdowns

Table 5.5 shows the cost breakdowns of the three optimized alignments generated from the proposed methodology with different design objectives.

Table 5.5 Cost Breakdown

	Alignment A1	Alignment A2	Alignment A3
Initial Cost (M)	43.6	105.4	45.4
Operation Cost (M)	86.6	72.9	77.3
User Cost (M)	60.1	54.0	55.8
Operation + User Cost (M)	146.6	126.9	133.1
Total Cost (M)	190.3	232.3	178.5

Among three alignments, Alignment A1 has the lowest initial costs of $43.6M, A2 has the lowest operation and user costs of $126.9M, and A3 has the lowest total costs of $178.5M.

Comparing Alignment A3 with Alignment A1, the initial costs increase by only $1.8M (4.1%), but the reduction in operation costs and user costs is $9.3M (10.7%) and $4.3M (7.2%), respectively. The total costs decrease by $11.8M (6.2%). Comparing Alignment A3 with Alignment A2, the operations costs and user costs increase by $6.2M (4.9%), but the reduction in initial costs is $60.0M (56.9%). The total costs decrease by $53.8M (23.2%). Thus, Alignment A3 demonstrates a better balance between low operation/user costs and low initial costs.

5.3 Summary

This chapter presents an extension that improves the base model by accounting for the impacts of vehicle dynamics on system operation cost and user cost in designing track alignments connecting several major stations. The proposed methodology can realistically simulate the movements of trains, in order to reliably estimate travel time and energy consumption, which are two of the most critical parameters in calculating operation costs and user costs. The generated alignments satisfy geometry constraints both along the alignment and at stations, and achieve the desirable trade-off between the initial cost invested and the operation/user cost incurred throughout the system's life cycle. The methodology proposed here can help designers to optimize rail transit alignments if given the major transit stations, or to initially evaluate different station locations. Such a tool can significantly speed up the design process and yield highly cost-effective solutions.

The method is tested in a hypothetical topography scenario to illustrate the essential trade-off among different system costs. A sensitivity analysis is presented to demonstrate the impacts of maximal operation speed, interest rate, unit price of energy cost and unit price of user cost on the design of rail transit alignments. The Baltimore Red Line study area is used in the case study. It has five stations pre-located based on the empirically designed alignment alternative. The numerical results demonstrate that the model can find very good solutions in regions with complex topographical features.

The proposed methodology is intended for designing rail transit alignments that connect several preset major stations. The demand used here for estimating user

and operational costs is assumed to be obtainable from external travel forecasts and to be insensitive to travel time. These assumptions, although widely used in real-world rail transit design practice, cannot effectively capture the interactions among rail transit demand, station location and alignment. The proposed method might be further improved by developing an integrated optimization model to concurrently optimize rail transit alignment and station locations. This integrated model should include additional decision variables for station locations and type, while relating the rail transit demand forecasts to the travel times and accessibility resulting from the jointly optimized alignments and station locations.

Chapter 6: Concurrent Optimization of Rail Transit Alignment and Station Locations

The proposed base model in Chapter 3 constitutes a practical rail transit alignment optimization methodology for generating alignments that pass through preset station locations. However, the determination of station locations may not always be straightforward. Planners sometimes have to identify the potential station locations and select the best set among these locations, while accounting for various geometric, topological, environmental and financial constraints. To address such a practical design scenario, this chapter presents the second extension of the base model, which aims to concurrently optimize station locations and the rail transit alignment connecting these stations.

Section 6.1 presents the system framework of the proposed concurrent optimization model, whereas its two critical modules are addressed in much detail in the following two sections. Section 6.2 explains how the proposed methodology generates the candidate pool of potential rail transit stations for the concurrent optimization model to choose from, and Section 6.3 presents the embedded rail transit demand forecasting module and its interaction with the optimization model. Section 6.4 presents a case study which demonstrates the applicability of the proposed concurrent optimization model and Section 6.5 summarizes the work and findings.

6.1 System Framework

The proposed model is designed to concurrently optimize rail transit alignment and station locations, and also help planners decide the type of each station. The system framework of the proposed concurrent optimization model of rail transit alignment and station locations is shown in Figure 6.1.
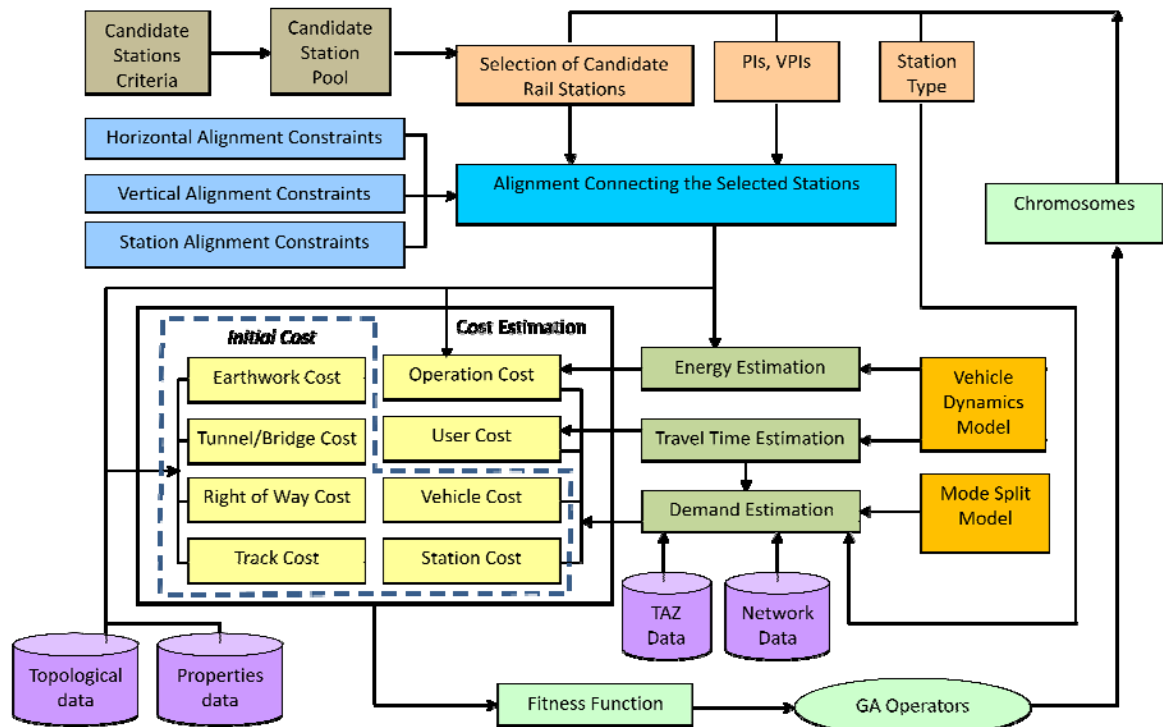
Figure 6.1 System Framework of the Concurrent Optimization Model

Compared to the framework in Chapter 5 for the first extension of the base model, which only incorporates vehicle dynamics, the concurrent optimization model addresses some additional modeling issues, as discussed below.

6.1.1 Additional Decision Variables for Station Locations and Station Type

To optimize station locations, this chapter will first develop a procedure for generating a candidate pool of station locations to satisfy various general requirements, such as topological features, accessibility to the existing roadway network, and land availability. Users can further refine these candidate stations via a graphical interface. The proposed concurrent optimization model will then select the best set of these candidate sites while generating the best alignment connecting the selected stations.

Both procedures for generating candidate station locations and the optimization model have to account for another important planning decision, i.e., the type of the rail transit station to be constructed. The two types of stations, pedestrian-orientated or park-and-ride facilities, target different group of users, have different attraction radii, involve different land needs, and require different costs. Generally, densely populated urban areas or suburban centers have a significant fraction of transit riders that access transit services on foot in order to avoid the costs associated with owning/driving/parking a vehicle, while suburban areas typically have many transit riders relying on access by auto. The proposed concurrent optimization model in this chapter will explicitly account for such differences in the modeling process, and let the optimization algorithm automatically select the type of each station.

6.1.2 Forecast Rail Transit Demand

The literature review in Chapter 2 indicates that one of the most important design criteria in existing station location models is the rail transit ridership these

stations can attract. Such ridership directly determines the rail transit system benefits in congestion relief, energy consumption and emission reduction, safety improvement, and user cost savings. As ridership forecasting is not the research focus of this dissertation, this chapter incorporates a discrete choice model, which is a widely-accepted transit ridership forecasting model in real-world practice, in the proposed concurrent optimization framework, and presents it in detail.

Ridership maximization and cost minimization, two of the most important objectives associated with the rail transit design problem, are generally conflicting in nature. For example, potential stations sites that could attract more ridership are usually located at more developed areas and thus associated with higher right-of-way cost. Shorter alignments between stations may lead to reduced travel time and are more appealing to travelers, but they may also require higher construction cost due to local topological features. This chapter will address how the forecasted ridership can be incorporated in the evaluation framework to more comprehensively evaluate different design alternatives.

6.2 Generation of the Candidate Pool of Potential Rail Transit Stations

The proposed concurrent optimization model will directly address the tradeoffs between ridership and cost in selecting station locations and generating the alignment in between. However, it would be impossible to check every point in the study network as potential station sites. This section will present procedures that apply quantified constraints to screen the study area and build a candidate pool of possible station locations. This candidate pool is then used as an input in the

integrated station and alignment optimization model to identify the best station locations.

Based on engineering practice, the candidate station locations typically need to satisfy the following general requirements:

1. Stations cannot be located within infeasible areas (e.g. lakes or rivers), environmental sensitive areas (e.g. wetland or residence of protected species), or historical sensitive areas (e.g. churches or cemetery).

2. Stations should have the potential to attract more ridership, which can be realized in three ways. First, the catchment areas of stations could cover a minimal amount of households or employment positions. The size of the catchment area, however, is related to the station type to construct: walking based stations have a shorter radius compared to park-and-ride stations. Secondly, stations could be located at existing activity centers or transfer centers of railway or bus transit systems. Finally, areas having the potential to support future growth at higher densities, such as centers of vacant land for future Transit Oriented Development, could also be good candidate station locations.

3. To attract more ridership, stations should have good accessibility for their target population. Park-and-ride stations should have easy access to the existing road network and preferably be near the roadways carrying significant traffic volumes; walking-based stations should have good accessibility for pedestrians.

4. Stations should avoid locations that could incur extremely high cost, such as extensively developed neighborhood with expensive right-of-way cost. The park-and-ride stations should not be too close to downtown; otherwise commuters would be highly unlikely to use rail transit which is considered to be a short trip.

Using these aforementioned principles, this section presents the following procedures to generate the candidate pool of possible rail transit stations.

Step 1: create a layer of grids inside the study area, Ω_s, with attributes:

S_w^i, feasibility of grid i to be pedestrian-orientated station (0 – infeasible; 1– feasible)

S_p^i, feasibility of grid i to be park-and-ride station (0 – infeasible; 1– feasible)

Set $S_w^i = 0$, $S_w^i = 0$

Step 2: create a layer Ω_i for infeasible areas, which combines:

- Wetlands

- Historic districts

- Historical sensitive area

- Topography features such as rivers, lakes, and valleys

Step 3: overlay the two layers from Step 1 and Step 2 to create the feasible grid layer

$$\Omega_f = \Omega_s \cap \overline{\Omega_i} \tag{6.1}$$

Step 4: Find all grids with the number of households within walking distance higher than a threshold

$$\text{For each } G_i \in \Omega_f, \text{ set } S_w^i = 1 \text{ if } \left(\sum_j H_j \,\middle|\, R_j^i \leq R_w \right) \geq H_w \qquad (6.2)$$

where H_j is the household number of block j, from census data

R_j^i is the distance from the center of census block j to G_i

R_w is the walking distance

H_w is the pre-specified threshold number of households within

walking distance

Step 5: Find all grids with the number of households within driving distance higher than a threshold

$$\text{For each } G_i \in \Omega_f, \text{ set } S_p^i = 1 \text{ if } \left(\sum_j H_j \,\middle|\, R_j^i \leq R_p \right) \geq H_p \qquad (6.3)$$

where R_p is the radius for park-and-ride stations

H_p is the pre-specified threshold number of households within driving

distance

Step 6: Find all grids with the number jobs within walking distance higher than a threshold

$$\text{For each } G_i \in \Omega_f, \text{ set } S_w^i = 1 \text{ if } \left(\sum_j E_j \,\middle|\, R_j^i \leq R_w \right) \geq E_w \qquad (6.4)$$

where E_j is the employment number of block j, from census data

E_w is the pre-specified threshold number of jobs within walking

distance

Step 7: Find the grids close to rail transfer or bus stations

For each $G_i \in \Omega_f$,

set $S_w^i = 1$ if count of rail or bus stations which satisfy $D_j^i \leq R_w$ is greater

than N_s,

where D_j^i is the distance from the bus/rail station j to G_i

N_s is pre-specified threshold number of bus/rail stations

Step 8: Find the grids without extremely high ROW cost

For each $G_i \in \Omega_f$, set $S_w^i = 0$ if $\left(\sum_j C_j \left| L_j^i \leq L_w \right. \right) \geq C_w$ (6.5)

$$\text{set } S_p^i = 0 \; if \; \left(\sum_j C_j \left| L_j^i \leq L_p \right. \right) \geq C_p \qquad (6.6)$$

where C_j is the ROW cost for property j

L_j^i is the distance from property j to G_i

L_w is the impact distance for pedestrian-orientated station

C_w is the maximal allowed ROW cost for pedestrian-orientated station

L_p is the impact distance for park-and-ride station

C_p is the maximal allowed ROW cost for park-and-ride station

Step 9: Obtain the layer of annual average daily traffic (AADT) polyline features and

select those with AADT greater than a user input value to form a new layer Ω_A. Find

grids that are far away from these features.

For each $G_i \in \Omega_f$, set $S_p^i = 0 \; if \; B_i \cap \Omega_A = \emptyset$

where B_i is a buffer area around grid i with a buffer radius R_A

Step 10: Find grids that are close to downtown center

For each $G_i \in \Omega_f$ Set $S_p^i = 0$ if $K_i \leq K_d$

where K_i is the distance from grid i to downtown center

K_d is a pre-specified threshold distance from the downtown center

Step 11: Let $\Omega_s = \{G_i \in \Omega_f | S_w^i + S_p^i > 0\}$ be the candidate pool. Sort Ω_s, base on

the distance from the starting terminal $\Omega_s = \{G_1, G_1, \dots, G_i, G_{i+1}, \dots, |L_i^0 \leq L_{i+1}^0\}$,

where L_i^0 is the distance from the starting terminal to G_i

After the proposed procedures screen the study area and generate the candidate pool of potential station locations, the concurrent optimization model will encode the decision variable for the selection of potential station site G_i as X_i, where

$X_i = 1, if$ G_i is selected

$X_i = 0, if$ G_i is not selected

The selection of stations needs to satisfy the following constraints

a) Minimum number of stations N_L

$$\sum_i X_i \geq N_L \tag{6.7}$$

b) Maximum number of stations N_U

$$\sum_i X_i \leq N_U \tag{6.8}$$

c) Minimum spacing between any two selected stations

$$\forall X_i = 1, X_j = 1, i \neq j : Y_{i,j} \geq Y_{min} \tag{6.9}$$

where $Y_{i,j}$ is the distance along the alignment from Station i to Station j

Y_{min} is the minimum spacing required between stations

154

d) Maximum spacing between any two selected stations

$$\forall X_i = 1, X_j = 1, i \neq j : \ Y_{i,j} \leq Y_{max} \qquad (6.10)$$

where Y_{max} is the maximum spacing required between stations

f) Minimum distance to depart from the starting terminal

$$\forall X_i = 1, X_j = 1, j > i: \qquad L_j^0 - L_i^0 \geq L_0 \qquad (6.11)$$

where L_i^0 is the distance from the starting terminal to Station i

L_0 is the minimum spacing required to depart from the starting terminal

e) Minimum distance to approach the end terminal

$$\forall X_i = 1, X_j = 1, j > i: \qquad L_i^1 - L_j^1 \geq L_1 \qquad (6.12)$$

where L_i^1 is the distance from Station i to the end terminal

L_1 is the minimum spacing required to approach the end terminal

With these additional decision variables and constraints, the GA-based solution heuristic developed in Chapter 4 can be applied to generate alignments to connect the selected stations. The next section will present the model that forecasts the rail transit demands and the incorporation of rail transit demands in the fitness function for alignment evaluation.

6.3 Forecast of Rail Transit Demand

Existing station location models represented ridership attraction either by the number of rail transit users calculated with simple mode choice models, or by the alignment coverage estimated as line coverage or station coverage [Laporte et al., 2000]. Such methods are quite simplified compared to various transit ridership forecasting models that are used in rail transit planning studies, as reviewed in Section 2.2.2. As transit ridership forecasting is not the research focus of this dissertation, the discrete choice model, a widely-accepted transit ridership forecasting model in the real world practice, is incorporated in the proposed concurrent optimization framework and presented in detail.

6.3.1 Choice Modeling for Rail Travel Demand Forecast

Discrete choice models model the travelers' choice among different transportation modes. The choice modeling is based on the random utility theory, which assumes that the decisions maker's preference for a discrete alternative is captured by a value called a utility, and his/her choice is reflected in the choice set with the highest utility. Choice models can be aggregate or disaggregate, according to the type of input data. The aggregate approach directly models the aggregate share of all decision makers choosing each alternative as a function of the characteristics of the alternatives and socio-demographic attributes of the group. The disaggregate approach recognizes that aggregate behavior is the result of numerous individual decisions and to model individual choice responses as a function of the characteristics of the alternatives available to and socio-demographic attributes of each individual.

This chapter assumes the total trips matrix is known from external regional demand forecast models, and thus employs the aggregate choice models. The models use the trip matrix as input and split the matrix into separated matrices, one for each mode.

Depending on the logit structure for the alternatives in the study area, the proposed concurrent optimization model in this chapter employs two types of choice models in its rail ridership forecasting module: a multinomial logit choice model for pedestrian-oriented stations and a nested logit choice model for Park-and-Ride facilities.

Multinomial Logit Choice Model

The multinomial logit choice (MNL) model is the most widely used discrete choice model, as its formula for the choice probabilities has a closed form and is readily interpretable. MNL relies on the assumption of independence of irrelevant alternatives (IIA). The basic utility U_m for choosing alternative m in MNL model is:

$$U_m = V_m + \varepsilon_m \tag{6.13}$$

where V_m is the representation of utility using observed variables

ε_m is the unkown part which is treated as random

The MNL model is obtained by assuming that each ε_m is an independently identically distributed extreme value. The relation of the logit probability to representative utility is sigmoid, or S-shaped. This shape has implications for the impact of changes in explanatory variables. If the representative utility of an alternative is very low or high compared with other alternatives, a small change in the utility of the alternative has little effect on the probability of its being chosen. The

point at which the increase in representative utility has the greatest effect on the probability of its being chosen is when the probability is close to 0.5, meaning a 50–50 chance of the alternative being chosen. In this case, a small improvement tips the balance in people's choices, inducing a large change in probability.

For pedestrian-oriented stations, the structure of the MNL model is shown in the Figure 6.2.
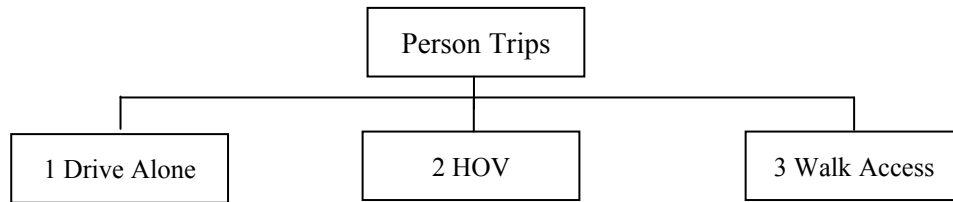
Figure 6.2 Multinomial Logit Choice Model

The probability of taking mode m between OD pair ij is given as:

$$P_{ijm} = \frac{e^{V_{ijm}}}{\sum_m e^{V_{ijm}}}$$

(6.14)

Here U_{ijm} is the utility of mode m between OD pair ij for a representative traveler. Representative utility is usually specified to be linear in parameters $V_{ijm} = \beta' x_{ijm}$, where x_{ijm} is a vector of observed variables relating to alternative m. With this specification, the logit probabilities become

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

(6.15)

Nested Logit Choice Model

The nested logit model (NLM), also known as generalized extreme value (GEV) model, allows partial relaxation of IIA property. It is useful when the unobserved portions of utility for some alternatives are correlated and IIA does not hold. A NLM is considered when the set of alternatives can be partitioned into subsets, called nests, so that the following properties hold:

1. For any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes or existence of all other alternatives. That is, IIA holds within each nest.

2. For any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests. IIA does not hold in general for alternatives in different nests

In the nested logit model, the utility is expressed as:

$$U_m = W_k + Y_m + \varepsilon_m \tag{6.16}$$

$$V_m = W_k + Y_m \tag{6.17}$$

Here the observed component of utility can be decomposed into two parts. The part W_k is constant for all alternatives within a nest and depends only on variables that describe nest k. These variables differ over nests but not over alternatives within each nest. The part Y_m depends on variables that describe alternative m and varies over alternatives within a nest k.

For park-and-ride stations, the structure of the nested logit model is shown in figure 6.3.
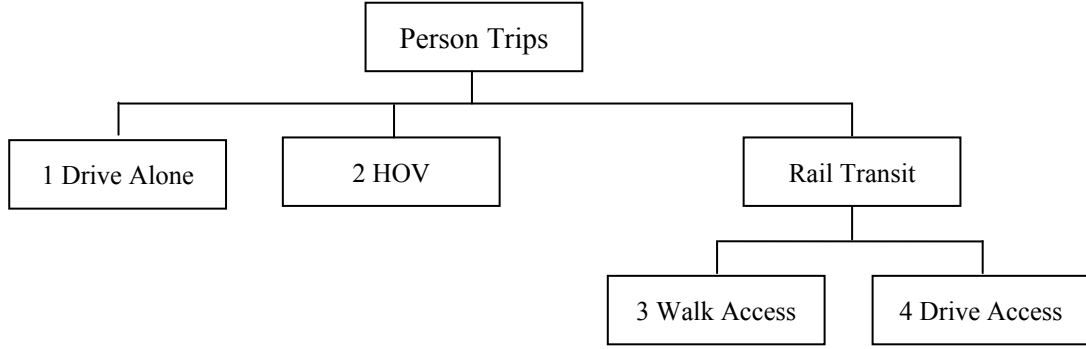
Figure 6.3 Nested Logit Model

The probability of taking mode m between OD pair ij is given as the product of two standard logit probabilities. The probability of choosing alternative $m \in B_k$, P_{ijm}, is the product of two probabilities:

- The probability that an alternative within nest B_k is chosen, P_{ijB_k}, which is the marginal probability of choosing an alternative in nest B_k

- The probability that then alternative m is chosen given that an alternative within B_k is chosen, $P_{ijm \mid B_k}$, which can be obtained by using MNL model

$$P_{ijm} = P_{ijm \mid B_k} \times P_{ijB_k} \tag{6.18}$$

$$P_{ijm \mid B_k} = \frac{e^{Y_{ijm}/\lambda_k}}{\sum_{l \in B_k} e^{Y_{ijl}/\lambda_k}} \tag{6.19}$$

$$P_{ijB_k} = \frac{e^{W_k + \lambda_k I_{ijk}}}{\sum_{l=1}^{K} e^{W_l + \lambda_l I_{ijl}}} \tag{6.20}$$

$$I_{ijk} = \ln \left(\sum_{m \in B_k} e^{Y_{ijm}/\lambda_k} \right) \tag{6.21}$$

The choice of nest is a marginal probability, also called the upper model. The choice of alternative within the nest is a conditional probability, also called the lower model. The quantity I_{ijk}, which is called the inclusive value or inclusive utility of nest k, links the upper and lower models by bringing information from the lower model

160

into the upper model. The coefficient λ_k of I_{ijk} in the upper model is called the log-sum coefficient. It indicates the degree of independence among the unobserved portions of utility for alternatives in nest B_k. A lower λ_k indicates less independence (more correlation).

6.3.2 Impacts of Forecasted Rail Transit Ridership on Fitness Evaluation

The proposed concurrent optimization model uses the total cost C_{total} as the fitness function, which is a function of initial costs C_c, operation cost saving C_O, and user cost saving C_U.

$$C_{total} = C_c + C_O + C_U. \tag{6.22}$$

Compared to Chapter 5 that assumes preset station locations and given transit ridership, in this chapter the operation cost saving and user cost saving are calculated based on the rail transit station-to-station demands estimated in the ridership forecasting module. The station construction cost is also included within the initial costs.

6.3.2.1 Operation Cost Saving

As formulated in Chapter 5, the operating cost includes energy costs and other operation and maintenance cost. In each time period p, the number of train trips needed N_p is calculated based on the estimated rail transit ridership.

$$N_p = \left\lceil \frac{\sum_i \sum_j D_{ijp}}{N_c D_c} \right\rceil \tag{6.23}$$

where D_c is the average number of passengers a train car can carry.

N_c is the number of cars per train.

161

Assuming the train service is only provided on workdays, the annual energy costs are:

$$A_e = (52 \times 5) \sum_p N_p \times E_R \times \frac{K_e}{10^6} \qquad (6.24)$$

where E_R is the round trip energy consumption (kwh)

K_e is the unit cost of energy ($/kwh)

The railway operation and maintenance costs are:

$$A_m = (52 \times 5) \sum_p \sum_i \sum_j D_{ijp} L_{i,j} \times \frac{K_O}{10^6} \qquad (6.25)$$

where $L_{i,j}$ is the travel distance from station i to station j (mile)

K_O is the unit operation and maintenance cost for rail ($/passenger-mile)

The auto operation cost for the park-and-ride trips is:

$$A_P = (52 \times 5) \sum_p \sum_i \sum_j D^P_{ijp} L^P_{i,j} \times \frac{K_a}{10^6} \qquad (6.26)$$

where D^P_{ijp} is the number of park-and-ride trips from TAZ i to TAZ j in period p

$L^P_{i,j}$ is the auto travel distance for D^P_{ijp} (mile)

K_a is the unit operation cost for auto ($/passenger-mile)

The original auto operation cost for the rail riders:

$$A_a = (52 \times 5) \sum_p \sum_i \sum_j D^r_{ijp} L^a_{i,j} \times \frac{K_a}{10^6} \qquad (6.27)$$

where D^r_{ijp} is the number of trips from TAZ i to TAZ j using rail in period p

$L_{i,j}^a$ is the auto travel distance for trip from TAZ i to TAZ j (mile)

The annual operating cost saving is:

$$A_o = (A_e + A_m + A_p) - A_a \qquad (6.28)$$

Assuming an annual interest rate of r and a life cycle of n_a years, the present value of operating cost saving over the life cycle is:

$$C_o = A_o \frac{(1+r)^{n_a} - 1}{r(1+r)^{n_a}} \qquad (6.29)$$

6.3.2.2 User Cost Saving

Similarly to the calculation of the operation cost saving, the annual user cost saving for railway riders is:

$$A_U = (52 \times 5) \sum_p \sum_i \sum_j D^r_{ijp} \times (T^r_{i,j,p} \times \frac{1}{3600} \times \frac{K^r_U}{10^6} - T^a_{i,j,p} \times \frac{1}{3600} \times \frac{K^a_U}{10^6}) \qquad (6.30)$$

where

$T^a_{i,j,p}$ is the travel time by auto from TAZ i to TAZ j in time period p (s)

$T^r_{i,j,p}$ is the travel time by rail from TAZ i to TAZ j in time period p (s)

K^a_U is the unit user cost for auto ($/passenger-hour)

K^r_U is the unit user cost for rail ($/passenger-hour)

The present value of user cost saving is:

$$C_U = A_U \frac{(1+r)^{n_a} - 1}{r(1+r)^{n_a}} \qquad (6.31)$$

6.3.2.3 Station Costs

The proposed concurrent optimization model assumes that station cost includes two parts: a fixed station cost that is independent of station locations, and a location-based station cost.

The fixed station cost includes the cost for station facility and the cost for parking facility. Assume the station site to be a rectangle shape with user specified length and width, the fixed cost for station facility varies only with the construction type, i.e., at-grade, elevated, or underground, which depends on the elevation difference between the proposed station and the existing ground. Assuming the cost of parking facility is linear with respect to the park-and-ride demands, the fixed cost for parking facility is calculated based on a preset unit cost per parking space.

The location-based station cost includes the ROW cost and the earthwork cost. Knowing the shapes for the station facility and parking facility, the ROW impact area obtained in Chapter 3 is updated with the station and parking sites. The properties inside the updated ROW impact area will have the ROW costs for both alignment and stations. The earthwork cost for at-grade stations and parking facilities can be obtained via GIS.

6.4 Case Study

Using the real-world Baltimore City as the study area, this section aims to illustrate the data preparation procedures of the proposed concurrent station location and alignment optimization model, and to demonstrate its effectiveness compared to

the sequential optimization methodology where stations are first selected and alignment is then designed between these selected stations.

6.4.1 Data Preparation for the Proposed Concurrent Optimization Model

6.4.1.1 Data for Candidate Stations

Following the procedures presented in Section 6.2, this section first created a grid layer inside the study area with user specified grid size (1000 feet by 1000 feet in the case study). Then a series of GIS operations are applied to identify the grids for candidate pedestrian-oriented stations and park-and-ride stations, as shown below.

- Land use pattern

Certain types of land use will be excluded for railway alignment and stations, such as forest, river, wetlands, historical area, and some restricted area due to political or economic concerns. The grid layer generated from the previous step is overlay with the land use layer in GIS. All the grids intersected with those restricted zones are identified as infeasible grids for railway stations.

- Census Block data

Census data are obtained from United States Census Bureau. Year 2000 census data for Baltimore City and Baltimore County are used in the case study.
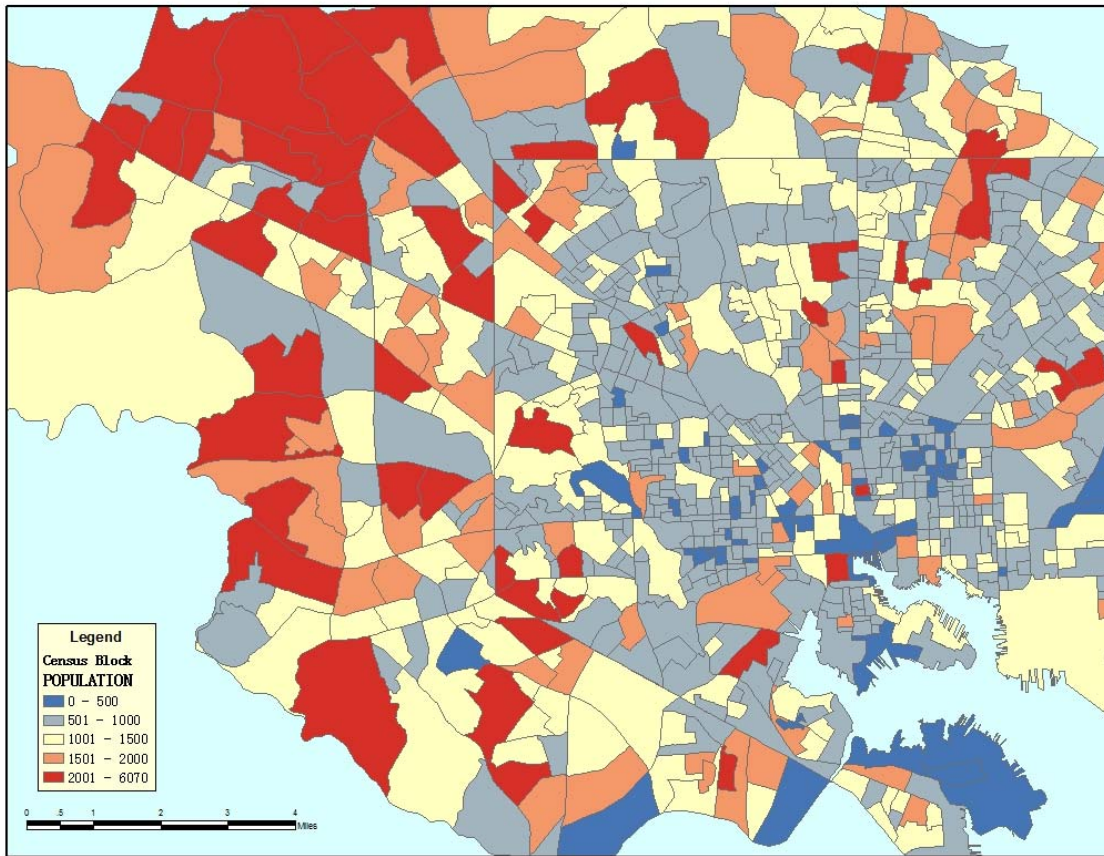
Figure 6.4 Census Block Data

The stations grids which attract high population or high employments are considered as potential station locations. Pedestrian-oriented stations and park-and-ride stations have different catchment area radius and thresholds of population and employments.

- AADT

Candidate park-and-ride stations need to meet the requirement of easy access to the existing road network. AADT line information obtained from Maryland State Highway Administration is used to determine the accessibility of potential park-and-ride station locations: the roadways near the candidate location should carry significant traffic volumes.

166

Figure 6.5 AADT Data

● Properties Data

Based on the properties distribution, a 250 feet by 250 feet grid layer is created in the study area with a ROW cost value for each grid. The candidate station locations should avoid the high ROW cost grids.

After applying the proposed procedures for generating the candidate pool of potential rail transit stations in Section 6.2, 52 candidate pedestrian-orientated station locations and 10 candidate park-and-ride station locations are found, as shown in Figure 6.6.
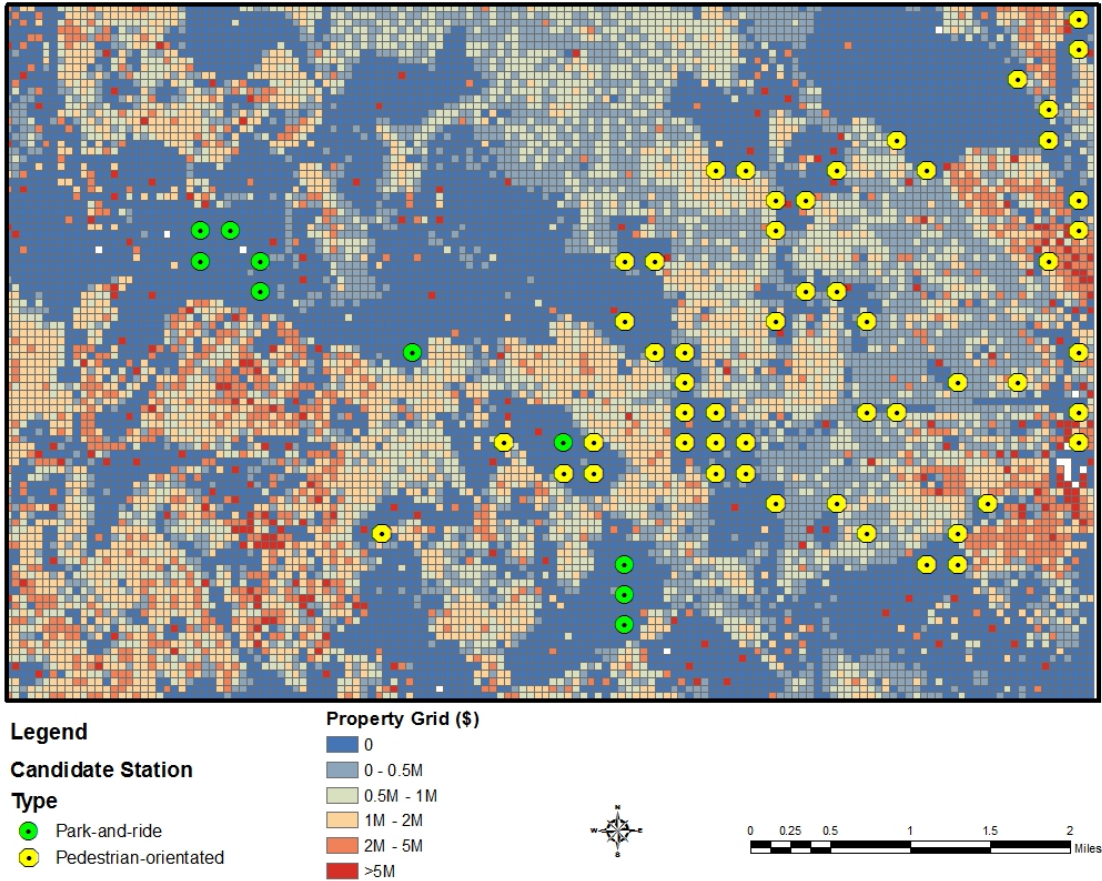
Figure 6.6 Candidate Station Locations

6.4.1.2 Data for Estimating Railway Travel Demand

As described in Section 6.3, the proposed concurrent optimization model applies a nested logit mode choice model to calculate mode choice for personal trips and output the following four trip tables:

- Mode 1: Drive Alone

- Mode 2: High Occupant Vehicle

- Mode 3: Walk to Rail

- Mode 4: Drive to Rail

The utility U_{ijm} is a function of the alternative characteristics and decision maker's characteristics, which include the following variables:

- T_{ijm}^{IN} − Travel time in the vehicle or train from TAZ i to j for mode m;

- T_{ijm}^{OUT} − Travel time outside of the vehicle or train from TAZ i to j for mode m;

- T^{w} − Waiting time or headway of the train at the boarding station

- C_{ijm} − Cost of mode m (gas, parking, and ticket) from TAZ i to j;

- S_{ijm} − Travel distance from TAZ i to j for mode m

- A_i − The number of autos per person in TAZ i

- B_j − Binary variable to check if the TAZ j is close to CBD

- E_j − Employment density of the TAZ j

This study considers all the Traffic Analysis Zones (TAZs) within 1 mile of candidate Pedestrian-oriented stations and/or within 5 miles of candidate Park-and-Ride stations. The TAZ data are obtained from Baltimore Metropolitan Council (BMC) models [Baltimore Metropolitan Council, 2004]. TAZ data contain zone related information, such as population, employment density, income, number of autos per person, and whether or not inside CBD.

The roadway network travel time information is also obtained from BMC models [Baltimore Metropolitan Council, 2004] for AM/PM peak and midday periods, which consider the congestion level for different time periods. All TAZ centers and candidate station locations are connected to the existing roadway network

via artificial connectors. TAZ to TAZ, TAZ to Station, and Station to TAZ travel time and distance matrices are generated using GIS shortest path function prior to the mode choice process. The model developed in Chapter 5 is used to compute the travel time and distance by train between any two stations.

All of the above data are used to calculate the variables in the utility function. For mode m:

$$U_{ijm} = (a_m^0 + a_m^1 T_{ij2}^{IN} + a_m^2 T_{ij2}^{OUT} + a_m^3 T_{ij}^W + a_m^4 C_{ijm} + a_m^5 S_{ijm} +$$

$$a_m^6 A_i + a_m^7 B_j + a_m^8 E_j) \tag{6.32}$$

where a_m^0 is the constant for mode m

$a_m^i, i = 1\ to\ 8$, is the coefficient for the aforementioned 8 variables for mode m

The case study in this section adapted the following values for constants, coefficients, and correlations from BMC model [Baltimore Metropolitan Council, 2004]. It is noted that the original model used different sets of parameters depending on the trip type (home based, work based, other based, etc…) and income level (I, II, and III). For simplicity, this section only uses home based trip and level II parameters in the mode choice modeling, which should be sufficient to examine the effectiveness of the proposed concurrent station and alignment optimization model.

Table 6.1 Nested Logit Model Parameters

	Drive Alone	HOV	Walk Access	Drive Access
m	1	2	3	4
a_m^0		-1.05989	1.29430	-0.55102
a_m^1	-0.0338	-0.0338	-0.0125	-0.0125
a_m^2	-0.0443	-0.0443	-0.0443	-0.0443
a_m^3 if $T_{ij}^w \leq$ 7.5 min			-0.0291	-0.0291
a_m^3 if $T_{ij}^w > 7.5\ min$			-0.0186	-0.0186
a_m^4	-0.1430	-0.1430	-0.0529	-0.0529
a_m^5		0.0991	0.3038	0.3038
a_m^6		-2.1822	-4.7928	-4.7928
a_m^7		0.3393	0.3393	0.3393
a_m^8			0.00003	0.00003
λ_m	1.0000	1.0000	0.7274	

After applying the above mode choice model to all OD pairs of TAZs in the study area, the total trip matrices in three time periods are split into 4 modes: drive alone, HOV, walk to rail, and drive to rail. The trip matrices for the latter two modes are used to compute the rail transit station to station demands, and are incorporated into the fitness calculation in the GA process.

6.4.2 Model Results

To test the effectiveness of the proposed concurrent station location and alignment optimization model, this section examined two optimization methods as follows:

- Two-stage optimization: locate stations first to maximize the demand, then find the alignment to minimize the cost
- Concurrent optimization: concurrently optimize the station locations and alignment to minimize the system cost

6.4.2.1 Comparison of the Two Optimization Methods

Table 6.2 presents the optimization results of the two-stage optimization and concurrent optimization.

Table 6.2 Comparison of Two-stage Optimization and Concurrent Optimization

	Two-stage	Concurrent	Difference
Number of Stations	5	5	0%
Total Daily Passengers	4,879	3,814	-21.8%
Travel Time (minutes)	11.8	10.8	-8.5%
Total Length (miles)	6.7	5.8	-13.4%
Total Initial Cost (M)	74.3	54.1	-27.2%
Operation/User Cost Saving (M)	-32.5	-36.5	12.3%
Total Cost (M)	41.8	17.6	-57.9%

Compared to the two-stage optimization, the concurrent optimization significantly reduces the total cost from 41.8M to 17.6M. The passenger trips for the

rail line from the concurrent optimization are 21.8% less than the maximum passenger trips found in two-stage optimization. By compromising in the passenger attraction, concurrent optimization reduces the travel time from 11.8 minutes to 10.8 minutes, shortens the track length from 6.7 miles to 5.8 miles, decreases the initial cost from 74.3M to 54.1M, and also decreases the operation and user cost from -32.5M to -36.5M. The numerical results prove the advantage of concurrent optimization over the two-stage optimization.

Figure 6.7 presents the station locations and horizontal alignments generated from the two optimization methods. Both alignments have three intermediate stations. The first intermediate station is the same. For the second and the third intermediate stations, the two-step optimization selected two dispersed locations to attract more railway passengers, whereas the concurrent optimization selected two closer locations to shorten the alignment length and travel time, so as to decrease the system total cost. The alignments are similar at both ends for the two optimization methods, whereas the middle pieces of the alignments are shifted to connect different stations selected.
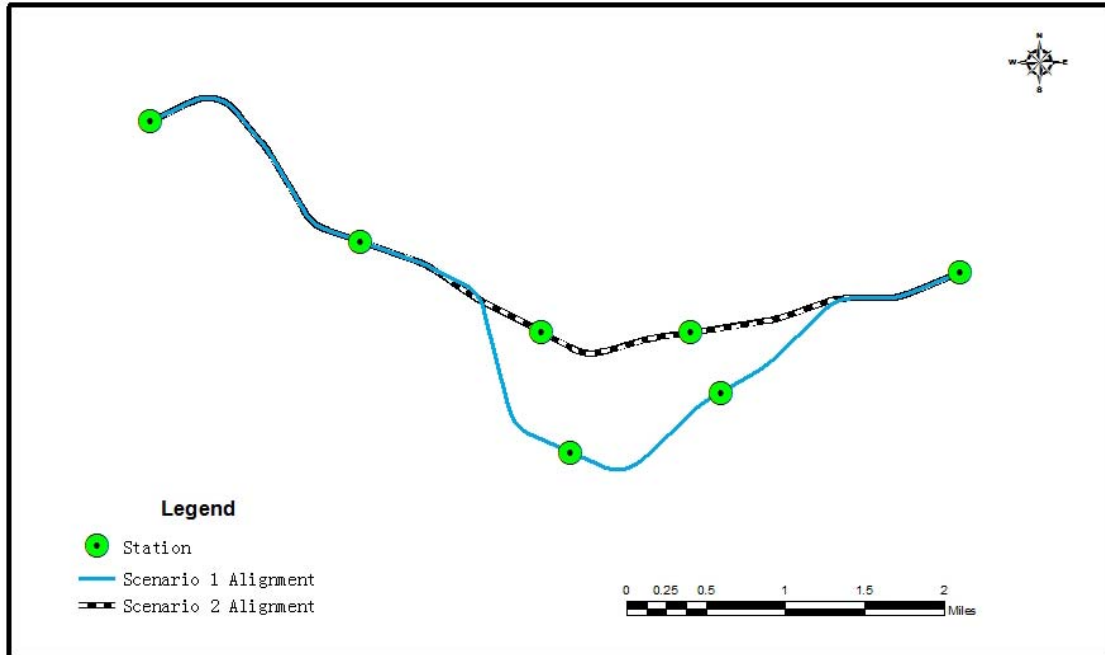
Figure 6.7 Optimized Station Locations and Alignments

6.4.2.2 Impact of Demand Variation on Optimization Results

This section aims to examine how the proposed concurrent optimization model adjusts its station selection and alignment design with variations in demand distribution so as to minimize the total cost. The design scenario adjusts the total demands from/to the four TAZs of 86, 87, 88 and 90 from 12,869 to 51,476, as shown in Figure 6.8. Figure 6.8 also compares the optimized station locations and alignments for the original and the adjusted demand distributions.
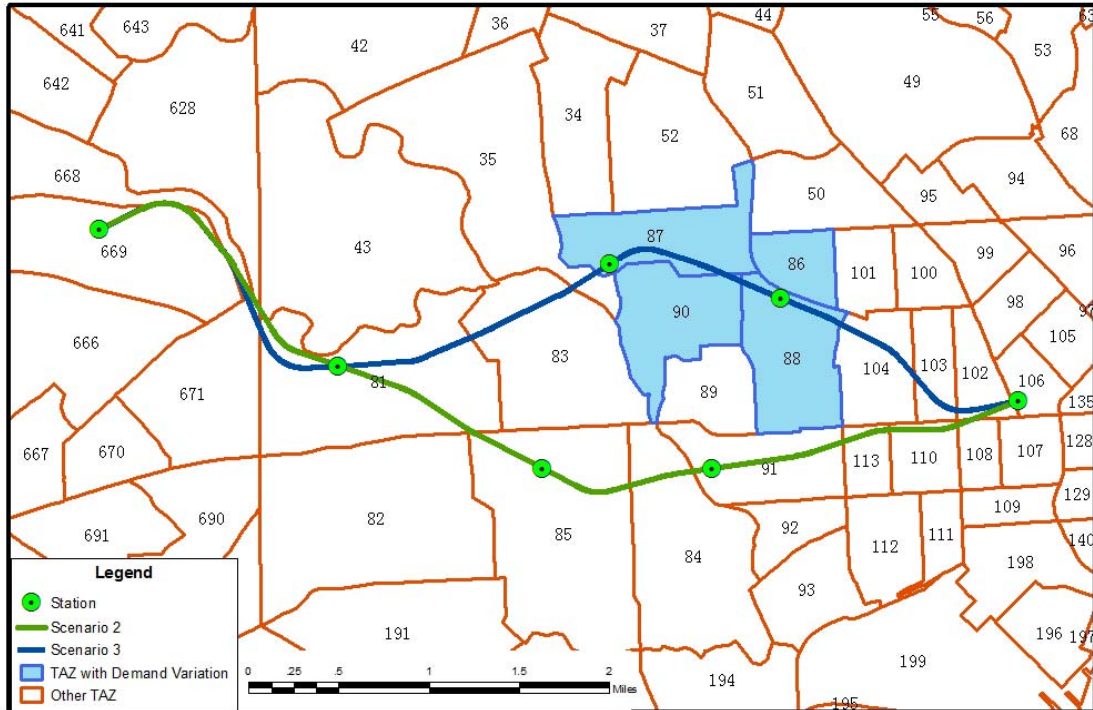
Figure 6.8 Impact of Demand Variation on Optimization Results

Two of the three stations selected for the original demand distribution are shifted to locate within the four TAZs with the adjusted demands. The two alignments start with the same segments at the western end of the study area, until they approach the first intermediate station. The solution algorithm then generates different tangent segments through the first intermediate station to adjust the alignment towards the two shifted station locations. Compared to the optimized station locations and alignment generated for the original demand distribution, the shifted station locations and alignment incurred an increase in the initial cost of 7.7M from 54.1M to 61.8M. However, the operation and user cost decreased 91.4M from -36.5M to -127.9M for the shifted station locations, as they attract more than twice of the original demand by directly serving the TAZs with higher demand. The results show that the algorithm is

175

effective in recognizing the demand patterns and can concurrently optimize the station locations and alignment accordingly.

6.4.2.3 Statistical Test of Solution Goodness

We apply a statistical method from Jong and Schonfeld [2003] to test the effectiveness of the proposed algorithm. This procedure is a sampling process. 50,000 random solutions are generated in scenario 2 and 23.7% of them (11,870) are feasible. The average cost of the feasible solutions is 161.1M, and standard deviation is 46.9M. The least cost of random solutions is 41.5M.

We use Gamma distribution and Normal distribution to fit the cost distribution from the feasible random solutions. The fitted Gamma distribution and Normal distribution are shown in Figure 6.9, with R^2 value 0.99 and 0.91 respectively.

The optimized cost from the proposed concurrent optimization model in scenario 2 is 17.6M, which is better than 99.89% of solutions in the fitted Normal distribution, and close to 100% of solutions in the fitted Gamma distribution. The result shows although we cannot guarantee a global optimal by using the proposed algorithm, the optimized solution is remarkably good compared to other solutions in the search space.
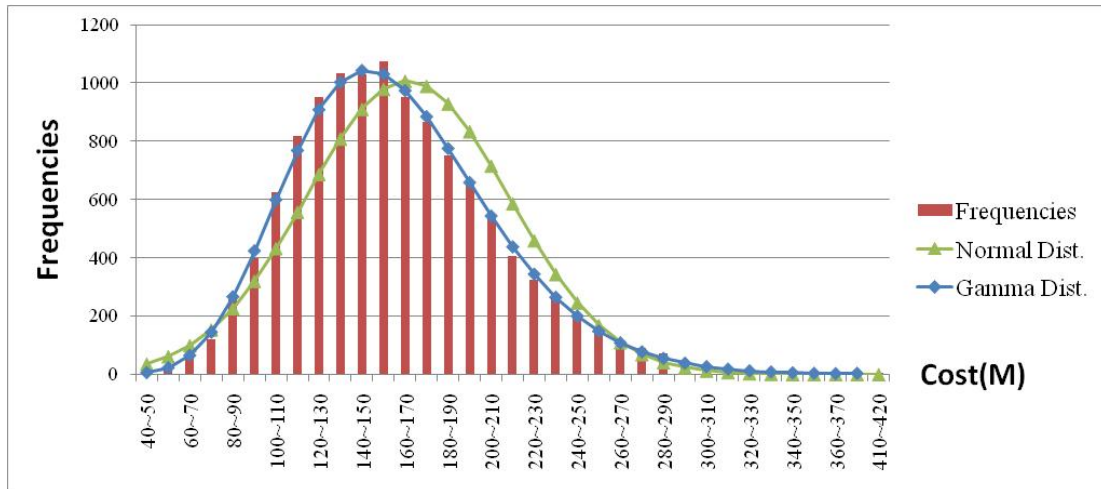
Figure 6.9 Statistical Test of Solution Goodness

6.5 Summary

This chapter presents a concurrent railway station location and alignment optimization methodology. The methodology first constructs the candidate pool of potential rail transit stations based on the comprehensive consideration of various requirements on topological features, accessibility to the existing roadway network, and land availability. These candidates are then selected along with the alignment in between using the concurrent optimization model to minimize the total cost of the system while satisfying station selection constrains. The total cost consists of initial cost, operation cost saving, and user cost saving. A nested logit choice model is applied for rail transit demand forecast to compute the operation cost saving and user cost saving.

The model accounts for the essential trade-off between cost minimization and ridership maximization in the total cost evaluation. On one hand, it may increase the

initial cost to accommodate more railway riders. On the other hand, it reduces operation and user costs by shifting more trips from auto to rail. This case study demonstrates the applicability of the proposed concurrent optimization model and its advantage over the traditional two-stage optimization.

Chapter 7: Optimizing Rail Transit Alignment

while Considering an Existing Road Network

The previous chapters presented a practical rail transit alignment optimization methodology for generating alignments that pass through preset station locations and two extensions for addressing the impacts of vehicle dynamics and transit demand on alignment design. This chapter presents a third extension, which aims to take into account the existing road network in concurrently optimizing station locations and rail transit alignments. In practice, most trail transit systems take advantage of the existing road network or abandoned railways to reduce costs, and the design process is usually empirical and very time-consuming. To address such design scenarios, the proposed optimization methodology has the capability to help planners identify the feasible links in the existing network that can accommodate the transit alignment, choose the best set among these links, and connect the selected links as well as the transit stations with viable alignments that meet various geometric, topological, environmental and financial constraints.

Section 7.1 presents the problem statement and identifies the issues for modeling existing network in rail transit design. Section 7.2 proposes criteria for identifying the feasible links of the existing network that rail alignment can use and the special GA operator designed to take advantage of these links in the optimization process. Section 7.3 presents a case study that demonstrates the effectiveness of the proposed optimization methodology. Section 7.4 summarizes the work and findings.

7.1 Problem Statement
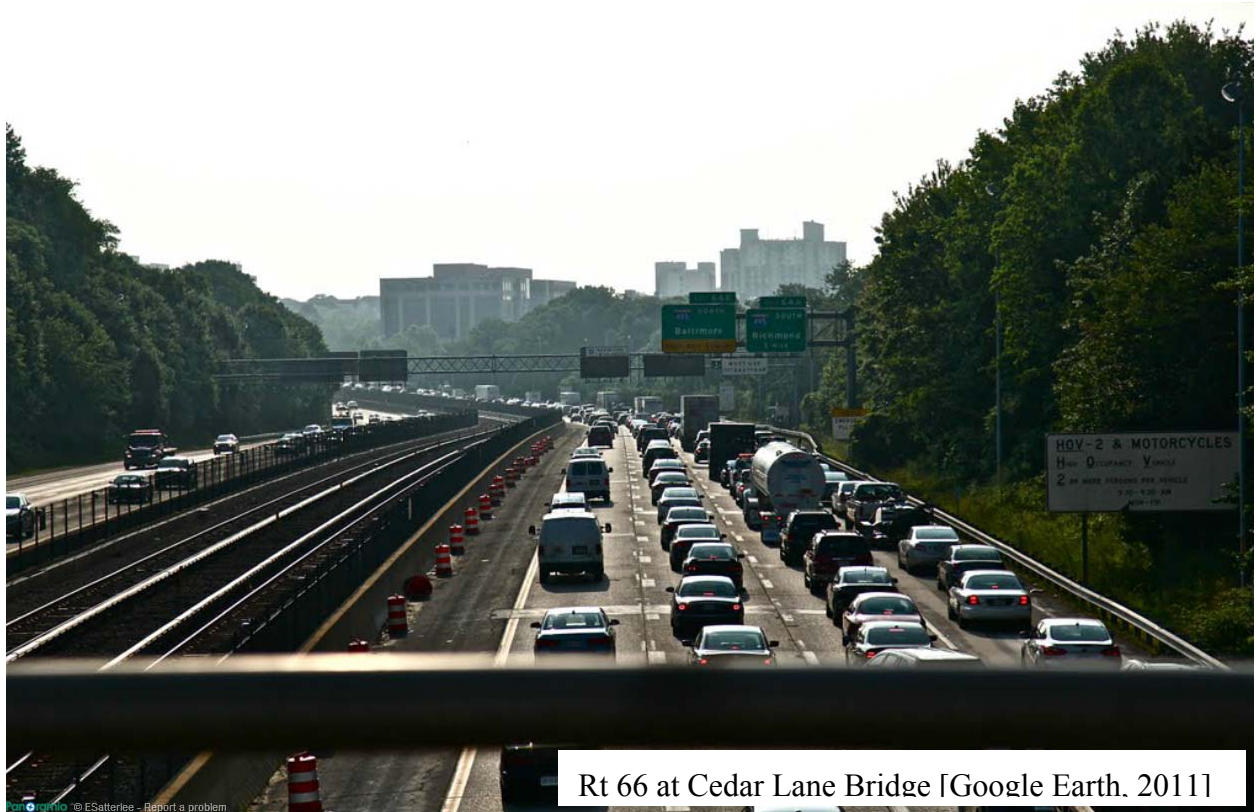
Figure 7.1 Illustration of a Typical Existing Roadway with Rail Transit

Figure 7.1 shows a typical section of a metro line in the Washington Metropolitan Area, which is in the median of an existing expressway I-66 from central Washington out to the western suburbs. This is a very common design practice in modern urban/suburban rail transit systems, where the entire alignment or at least a portion of it would follow the existing roadways or abandoned railway corridors. It can be particularly advantageous if the cross-section of the existing roadway is sufficient to accommodate the proposed rail transit line in its median or shoulder. This practice is cost-effective and of little impact to surrounding communities by

reducing right-of-way acquisition. None of the alignment optimization models in literature can account for such design practice in the optimization process.

The existing roadway or abandoned railway corridors typically have two notable impacts on rail transit design: On one hand, existing roadways may provide right of way in their medians or along their outside clear zones for the transit alignment to utilize and thus reduce construction cost; on the other hand, existing roadways may also create difficulties in crossing and access design.

To address these impacts, the proposed optimization methodology should have the following capabilities:

- Identify the feasible links of the existing network that have sufficient right of way to accommodate the transit alignment.

- Choose from feasible links of the existing network the rail transit alignment may follow. Special decision variables and genetic operators should be designed to avoid frequent entering/exiting the median along the same roadway.

- Generate viable alignments which connect the selected links as well as the transit stations. The algorithm should account for the geometric requirements on both the horizontal and vertical alignments.

- Estimate the construction/operation cost by accounting for the reduced right-of-way and the additional crossing/access when an existing link is utilized by the proposed alignment.

Given the two end terminals of the target rail transit line, the outputs of the proposed optimization methodology include:

- The location of intermediate rail transit stations;

- The horizontal alignment with an optimal combination of new rail lines and selected existing links, and the transition between them;

- The vertical alignment with the optimal profile along the new rail lines, its transition to existing profile along existing links and its transition at the crossings with the existing roadway network.

- The total cost and its break-down.

7.2 Modeling Procedures

This section first presents the procedures for generating candidate links from the existing network that can accommodate the proposed rail transit alignment. Then a special GA operator is designed to allow the alignment to follow the existing links. Finally, this section presents a comprehensive cost estimation framework to evaluate the generated alignments.

7.2.1 Select Candidate Links from the Existing Network

The candidate links should have sufficient cross-section to accommodate the proposed rail transit alignment, and can be one of the following:

- Abandoned railways;

- Existing roadways with wide shoulder ;

- Existing roadways with wide median ;

- Existing roadways with modifiable cross-section, where the median or shoulder can be widened by shifting the lanes, or where some lanes can be converted to rail. Higher costs are associated with these two scenarios.

The screening of an existing network based on the above criteria identifies the feasible links of the existing network that the proposed rail transit line can use. Then, for any pair of candidate links, if they can be connected solely by existing candidate links, a generic shortest path procedure is applied to generate candidate existing paths between them with the minimal total path cost. The path cost includes link cost, connector cost, and turning penalty.

The link cost $C_{i_p}^L$ is a given value and is decided by the location of link i and the potential position of rail alignment in its cross-section. The position indicator p is defined as: p=0 median; p=1 left shoulder; p=2 right shoulder.

The connector cost $C_{i_p j_q}^C$ is a function of the conflict traffic volumes for the movement from link i position p to link j position q.

The turning penalty $P_{i_p j_q}^C$ is applied for all connectors when the train needs to slow down to make turns. The turning penalty is higher with a sharper turn.

The path cost between two links m and n is the sum of costs for all links and connectors in between.

$$C_{mn}^P = \Sigma_{\text{link } i \in S_{mn}} C_{i_p}^L + \Sigma_{\text{connector } ij \in S_{mn}} (C_{i_p j_q}^C + P_{i_p j_q}^C) \qquad (7.1)$$

where S_{mn} is the shortest path between link m and n.

Set $Y_{mn}^P = 1$ if link m and link n can be connected solely by existing candidate links, 0 otherwise. Store this indicator in database, along with the minimal path cost if applicable, for any pair of candidate links for use later in Section 7.2.2.

7.2.2 GA Special Operator

The proposed optimization methodology introduced a decision variable to indicate whether or not a PI point can be relocated to an existing link by the GA special operator: $X_s^i = 1$ if PI i can be automatically relocated, and 0 otherwise.

With the additional decision variable, the proposed GA special operator applies the following rules:

1. A PI is assigned to an existing link if the following conditions are met:

    a) The PI has $X_s^i = 1$;

    b) The distance between PI and the existing link on the cutting plane is less than a threshold;

    c) The distance between PI and the existing link is shorter than all other candidate links.

2. If two PIs are assigned to the candidate links, the existing path connecting these two candidate links is retrieved from the database and used to connect the two PIs.

3. If a path is found between two PIs, all PIs between them are ignored.

Following the above rules, the GA special operator consists of the steps given below:

1. Obtain PIs from GA

2. Intersect existing network links with the cutting planes to obtain the following two parameters:

   a) P_j^i Intersection point of existing link i with cutting plane j

   b) $E_j^{i,1}, E_j^{i,2}$ extension points, which satisfy

      i. Line $(P_j^i, E_j^{i,k})$ tangent to link i, k=1,2

      ii. $\left\| P_j^i, E_j^{i,k} \right\| = L_e$, which is an user input extension length, k=1,2

3. Assign each PI to the nearest existing link

   $\forall PI_j$

   Set $N_j = i$, assign PI_j to the nearest existing link i. In this case, the following three constraints should be satisfied:

   $X_s^j = 1$ and

   $\left\| PI_j, P_j^i \right\| \leq \left\| PI_j, P_j^k \right\|$, where $i \neq k$, and

   $\left\| PI_j, P_j^i \right\| \leq L_p$, where L_p is an user input threshold

   Set $N_j = 0$, PI_j is not assigned to any existing link, when one of the following is true:

   $X_s^j = 0$ or

   $\left\| PI_j, P_j^k \right\| > L_p$, for any existing link i

4. Relocate PIs onto the existing network

   Between each pair of neighboring rail stations there are n cutting planes. The operator searches from both ends, until it finds two PIs which are both assigned to their nearest existing links and there is an existing network path connecting the two

links. These two PIs are relocated onto their assigned existing network links and all PIs between them are ignored.

Step 4.0: $j=0$

Step 4.1:   If $N_j=0$ then

   Go to step 4.5

Step 4.2: $k=n+1$

Step 4.3:   If $N_k>0$ and $Y_{jk}^P = 1$ then

   Set $PI_j = P_j^{N_j}$

   Set $PI_k = P_k^{N_k}$

   Add the path between $P_j^{N_j}$ and $P_k^{N_k}$ to the alignment

   Insert new $PI = E_j^{i,1}$ before PI_j

   Insert new $PI = E_k^{i,2}$ after PI_k

   $j=k+2$

   Go to step 4.6

   Otherwise

   $k-1 \rightarrow k$

Step 4.4:   If $k>j$ then

   Go to step 4.3

Step 4.5: $j+1 \rightarrow j$

Step 4.6:   If $j<n$ then

   Go to Step 4.1

Step 4.7:   End

5. Complete the alignment by adding rail segments connecting stations to paths, and also connecting between paths, applying the alignment methodology in Chapter 3

    The following figures illustrate two examples of possible scenarios using an existing network in the rail transit alignment. In the first scenario, there are 4 cutting planes between station 1 and station 2. Two stations are both located on existing network links (station 1 on Link 1 and station 2 on link 4). Path 1 connecting link 1 and link 4 is fully used as the alignment connecting these two stations. All 4 PIs between two stations are ignored in this scenario.
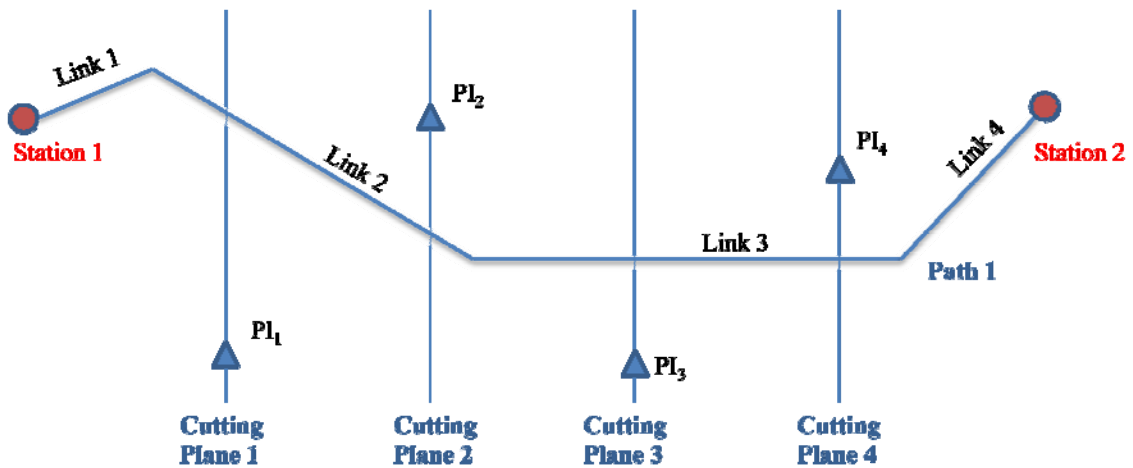
Figure 7.2 Use Existing Network Example 1

    In the second scenario, there are 6 cutting planes between station 1 and 2. There is no path directly connecting two stations since neither of them is located on an existing network link. The alignment in this scenario consists of two existing network paths and three new rail segments. The first path connects PI_1 and PI_3, and the second path connects PI_5 and PI_6. Two of the three new rail segments connect two stations to the paths, and rail segment 3 connects path 1 and path 2.
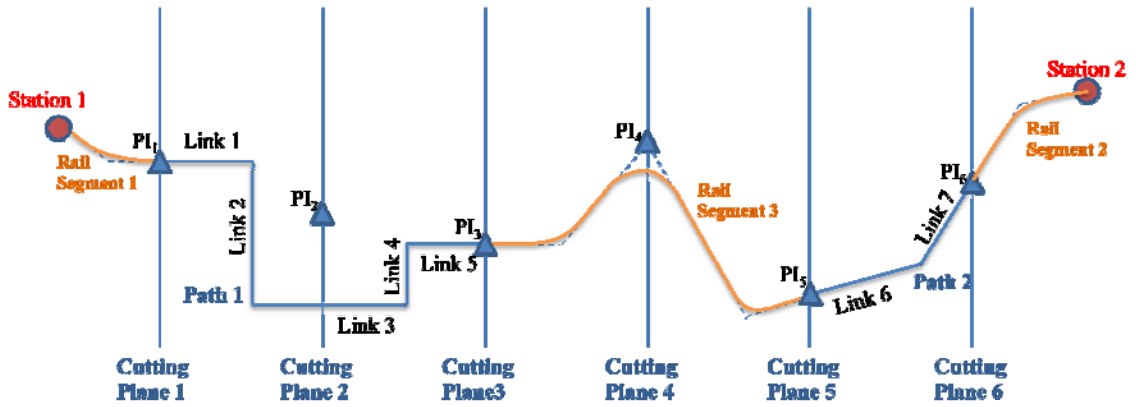
Figure 7.3 Use Existing Network Example 2

7.2.3 Cost Estimation:

As presented in Chapter 6, the total cost is defined as a function of initial costs C_c, operation cost saving C_O, and user cost saving C_U.

$$C_{total} = C_c + C_O + C_U. \tag{7.2}$$

7.2.3.1 Initial Costs

The initial cost C_c includes the cost of constructing new alignment segments C_c^s, the cost of using the existing network C_c^e, track cost C_L, train cost C_V, and station cost C_S.

$$C_c = C_c^s + C_c^e + C_L + C_V + C_S \tag{7.3}$$

Track cost C_L is determined by the overall track length and unit track cost. Vehicle cost C_V is decided by the train headway and run trip travel time.

The cost for constructing new alignment segments is the sum of earthwork cost, bridge and tunnel cost, right-of-way cost and track cost on all new alignment segments. These costs can be calculated by applying the methodology from Chapter 3.

$$C_c^s = C_E^s + C_B^s + C_T^s + C_R^s \tag{7.4}$$

The cost of using the existing network is the sum of costs of all existing links used in the alignment. The link cost is pre-determined for each link. It includes the guideway preparation cost which varies with the position of rail centerline at the cross-section, the right-of-way cost such as purchasing abandoned railway properties, and the cost for shifting existing roadway lanes if needed.

$$C_c^e = \sum_{i_p} C_{i_p}^L \tag{7.5}$$

The connector cost will be discussed in the user cost section.

7.2.3.2 Vehicle Dynamics on Existing Networks

The train is assumed to follow the speed limit on the existing roadway it travels. On connectors, a speed reduction zone is set up on all connectors with sharp turns. The vehicle dynamics are adjusted to reflect the speed change inside the existing network.

In each time interval, if the train travels on the existing network, the vehicle dynamics algorithm compares current speed with the speed limit on existing link or connector. Acceleration or deceleration is applied according to the speed difference. Grade information on existing links and turning radius on connectors are input to model the vehicle dynamics on existing network.

By adjusting the vehicle dynamics, the penalty of using slow existing links and making sharp turns on connectors are reflected in the energy consumption and travel time increase, which have significant impacts on the operation cost and user cost.

7.2.3.3 Operation Cost Saving and User Cost Saving

After the energy consumption and travel time are obtained, the operation cost saving and user cost saving can be calculated using the model presented in Chapter 6.

In calculating user cost savings, another cost is considered for the additional delay due to conflicts with surface traffic by introducing the rail movement through intersections.

$$C_{i_p j_q}^C = \frac{(1+r)^{n_a}-1}{r(1+r)^{n_a}} \times (52 \times 5) \times C_d \times \sum_{i_p j_q} d^C \times V_{i_p j_q}^C \qquad (7.6)$$

d^C is the average additional delay for the conflict traffic, which is a function of train headway and traffic control type. Preemption rail control can manipulate traffic signals in the path of the rail line, stopping conflicting traffic and allowing trains to pass without delays. Other control types, such as priority control, have less impact on conflict traffic, but the train must stop if it does not arrive during the train crossing phase. In the proposed model, preemption control is assumed for all connectors at rail line at-grade crossings. $V_{i_p j_q}^C$ is the conflict traffic volume of connector $i_p j_q$. C_d is the unit cost of delay.

7.3 Case Study

This section uses the same study area used in Section 6.4 for the Baltimore Red Line to demonstrate how the rail transit alignment is adjusted after accounting for the existing network in the alignment optimization. The study area covers a 7-mile east-west transit corridor in the western Baltimore City.

To illustrate the proposed methodology with different existing networks that are possible in practice, the case study assumes two different sets of existing networks

as input. The first network consists of 4 individual existing links without any connector, as shown in Figure 7.4. The second network has 15 links which are connected, as shown in Figure 7.5.
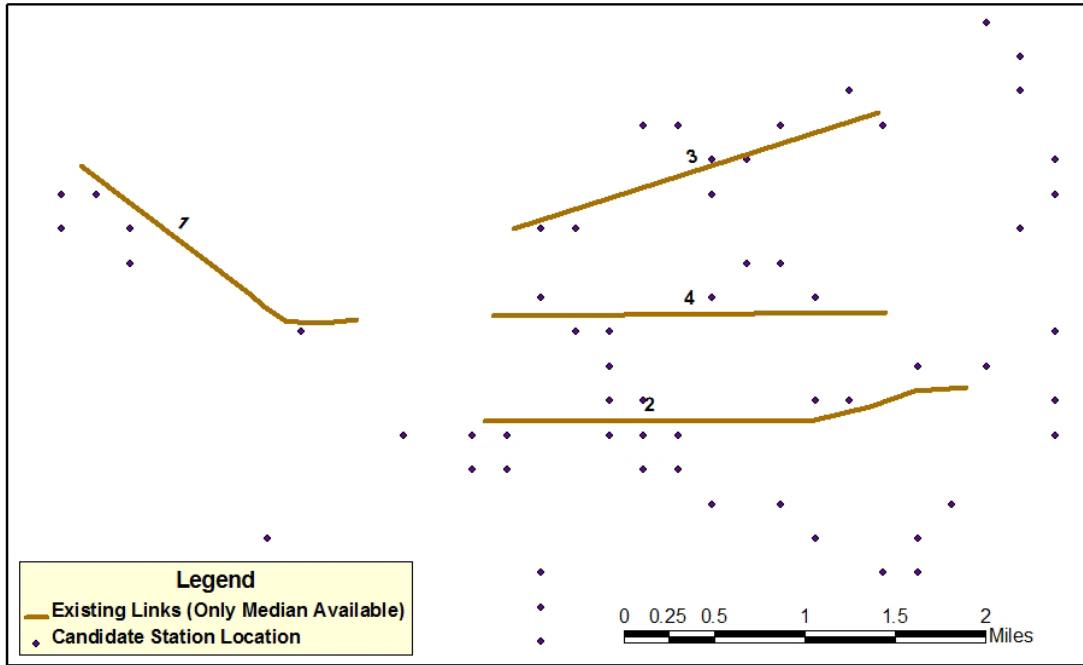
Figure 7.4 Existing Network 1 Map
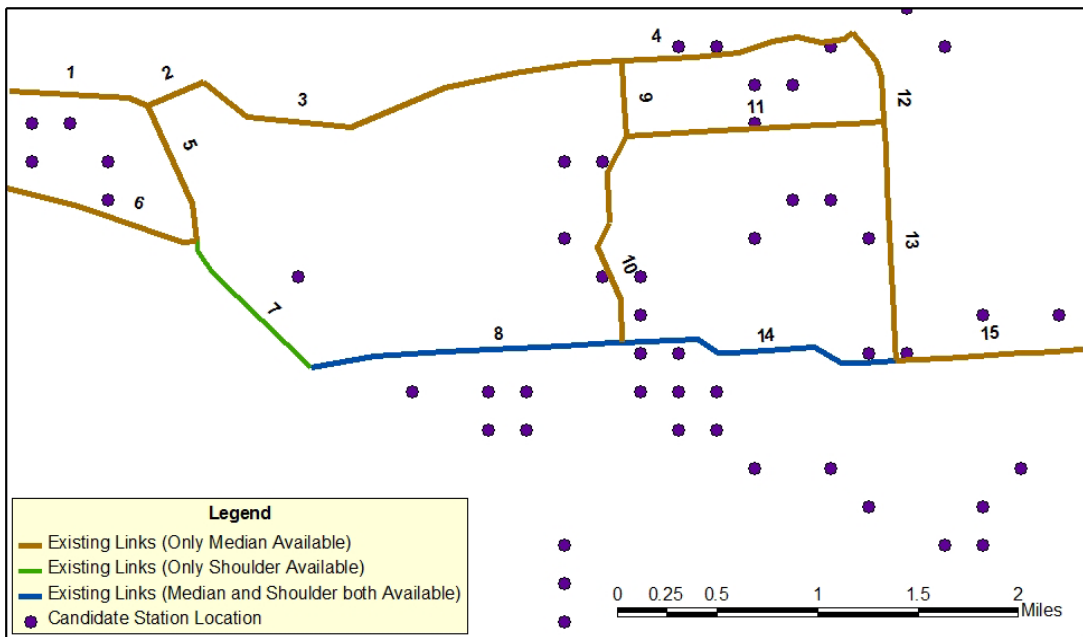
Figure 7.5 Existing Network 2 Map

7.3.1 Existing Network Related Datasets

The following Table 7.1 summarizes the link properties of the above two existing networks.

Table 7.1 Existing Network Link Properties

Network 1			Network 2		
Link	Position	Link Cost (M)	Link	Position	Link Cost (M)
1	Median	3.00	1	Median	0.17
2	Median	2.33	2	Median	0.08
3	Median	2.12	3	Median	0.55
4	Median	2.17	4	Median	0.29
			5	Median	0.18
			6	Median	0.34
			7	Shoulder	0.21
			8	Median	0.39
				Shoulder	1.56
			9	Median	0.09
			10	Median	0.27
			11	Median	0.32
			12	Median	0.12
			13	Median	0.27
			14	Median	0.35
				Shoulder	1.40
			15	Median	0.31

Network 1 does not have connectors. Connectors in network 2 are listed in the following Table 7.2. The turning penalty is calculated based on the reduced speed value on the connector and the delay cost is imposed on the traffic conflicting with the rail transit line. The connector cost is the sum of tuning penalty and delay cost.

Table 7.2 Existing Network 2 Connector Properties

From Link		To Link		Speed (mph)	Conflict Traffic	Cost (M)	From Link		To Link		Speed (mph)	Conflict Traffic	Cost (M)
1	Med	2	Med	30	1852	0.36	10	Med	9	Med	50	5808	0.83
1	Med	5	Med	30	2132	0.40	10	Med	11	Med	15	7181	1.52
2	Med	1	Med	30	1852	0.36	10	Med	14	Med	15	11074	2.08
2	Med	3	Med	20	2152	0.61	11	Med	9	Med	15	9588	1.87
2	Med	5	Med	15	1708	0.74	11	Med	10	Med	15	7181	1.52
3	Med	2	Med	20	2152	0.61	11	Med	12	Med	15	6508	1.43
3	Med	4	Med	50	7181	1.02	11	Med	13	Med	15	5262	1.25
3	Med	9	Med	15	2583	0.87	12	Med	4	Med	15	5189	1.24
4	Med	3	Med	50	7181	1.02	12	Med	11	Med	15	6508	1.43
4	Med	9	Med	15	3271	0.97	12	Med	13	Med	50	5808	0.83
4	Med	12	Med	15	5189	1.24	13	Med	11	Med	15	5262	1.25
5	Med	1	Med	30	2132	0.40	13	Med	12	Med	50	5808	0.83
5	Med	2	Med	15	1708	0.74	13	Med	14	Med	15	10950	2.06
5	Med	6	Med	10	6117	1.87	13	Med	15	Med	15	11737	2.17
5	Med	7	Shld2	50	1852	0.26	14	Med	8	Med	50	9588	1.37
6	Med	5	Med	10	6117	1.87	14	Med	10	Med	15	11074	2.08
6	Med	7	Shld2	30	7969	1.24	14	Med	13	Med	15	10950	2.06
7	Shld2	5	Med	50	1852	0.26	14	Med	15	Med	50	6508	0.93
7	Shld2	6	Med	30	7969	1.24	15	Med	13	Med	15	11737	2.17
7	Shld2	8	Med	30	11074	1.68	15	Med	14	Med	50	6508	0.93
8	Med	7	Shld2	30	11074	1.68	7	Shld2	8	Shld1	30	22148	3.26
8	Med	10	Med	15	11737	2.17	7	Shld2	8	Shld2	30	0	0.10
8	Med	14	Med	50	9588	1.37	8	Shld1	14	Shld1	50	0	0.00
9	Med	3	Med	15	2583	0.87	8	Shld2	14	Shld2	50	19176	2.73
9	Med	4	Med	15	3271	0.97	10	Med	14	Shld1	15	21325	3.54
9	Med	10	Med	50	5808	0.83	10	Med	14	Shld2	15	9588	1.87
9	Med	11	Med	15	9588	1.87	14	Shld1	15	Med	30	10950	1.66
10	Med	8	Med	15	11737	2.17	14	Shld2	15	Med	30	17458	2.59

Note: Med – Median; Shld1– EB/SB shoulder; Shld2 – WB/NB shoulder

After the links and connectors are defined, a generic shortest path algorithm is implemented to find out the path route and minimal cost for all pairs of existing network links. Network 1 has 4 unconnected links. Table 7.3 lists a part of the paths and their associated costs generated for the existing network 2.

Table 7.3 Existing Network 2 Paths

From Link	To Link	Path Cost (M)	Route
1	1	0.17	1
1	2	0.61	1 ~ 2
1	3	1.77	1 ~ 2 ~ 3
1	4	3.09	1 ~ 2 ~ 3 ~ 4
1	5	0.76	1 ~ 5
1	6	2.81	1 ~ 5 ~ 6
1	7	1.24	1 ~ 5 ~ 7
1	8	3.28	1 ~ 5 ~ 7 ~ 8
1	9	2.73	1 ~ 2 ~ 3 ~ 9
1	10	3.83	1 ~ 2 ~ 3 ~ 9 ~ 10
1	11	4.92	1 ~ 2 ~ 3 ~ 9 ~ 11
1	12	4.45	1 ~ 2 ~ 3 ~ 4 ~ 12
1	13	5.55	1 ~ 2 ~ 3 ~ 4 ~ 12 ~ 13
1	14	5.00	1 ~ 5 ~ 7 ~ 8 ~ 14
1	15	6.24	1 ~ 5 ~ 7 ~ 8 ~ 14 ~ 15
……			
15	14	1.59	15 ~ 14
15	15	0.31	15

7.3.2 Numerical Results

To test the effectiveness of the proposed alignment optimization model considering the existing network, this section examines two new alignments optimized with existing network 1 and existing network 2, and compares them with the alignment from the concurrent optimization in Chapter 6. The three alignments compared in this section are:

● Alignment 1: concurrently optimized station locations and alignment to minimize the system cost without considering the existing network

● Alignment 2: optimized station locations and alignment, considering the existing network 1

●   Alignment 3: optimized station locations and alignment, considering the

existing network 2

7.3.2.1 Horizontal Alignment Comparison

Figure 7.6 compares the horizontal alignments for alignment 1 and alignment

2.
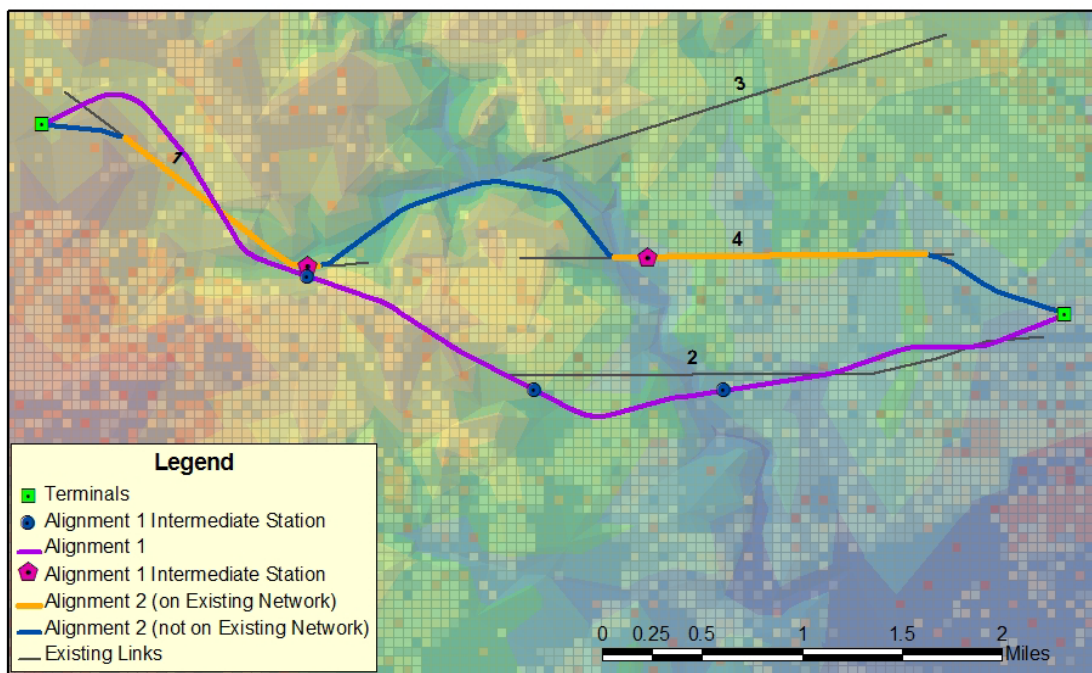
Figure 7.6 Horizontal Alignment Comparison between Alignment 1 and 2

Alignment 2 is assembled from three new alignment segments and two

existing network links. Two new alignment segments connect the terminals to the

existing network links No. 1 and No. 4 at both ends. The other new segment connects

these two existing links, and uses horizontal curves to avoid areas with high ROW

cost. The model also chooses the alignment with smaller grade changes to reduce the

earthwork cost. Two intermediate stations in alignment 2 are located on existing link No 1 and 4. The one on No 1 is the same station located in alignment 1 but shifted to the existing link.

Figure 7.7 compares the horizontal alignments for alignment 1 and alignment 3.
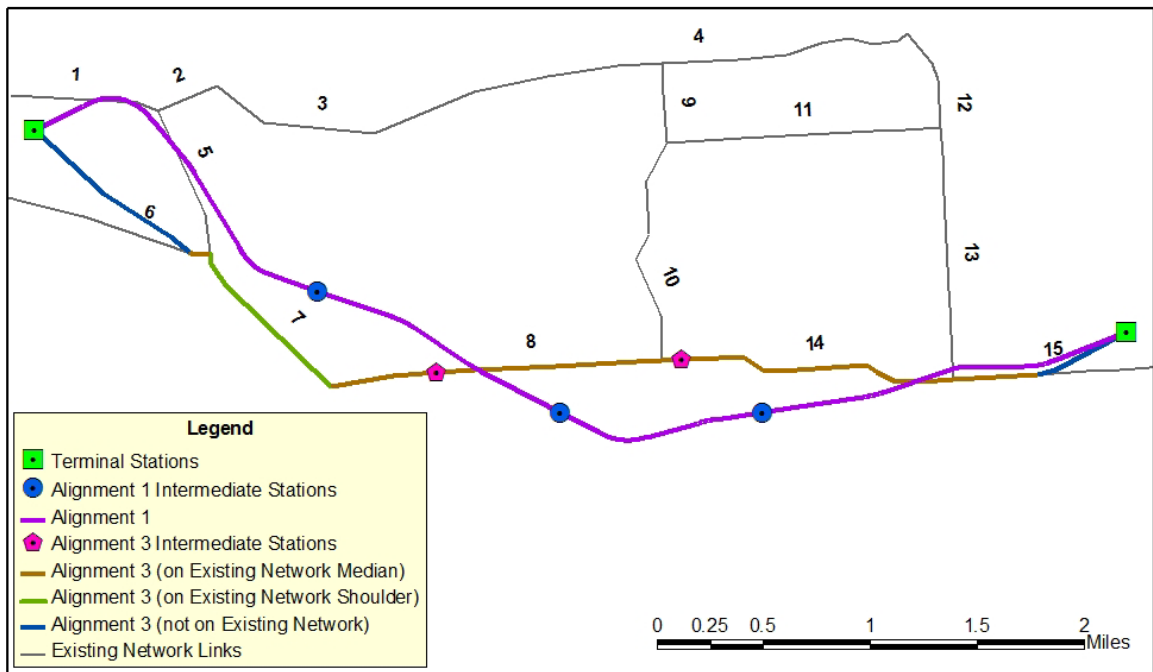
Figure 7.7 Horizontal Alignment Comparison between Alignment 1 and 3

Among 15 existing links, 5 are selected in alignment 3: No. 6, No. 7, No. 8, No. 14, and No. 15. These links determine a path with least path cost to connect link 6 and 15, which are connected to two terminals by new alignment segments at both ends. Alignment 3 uses the shoulder of link No. 7 and the median of other 4 links. Two intermediate stations in alignment 3 are located on links 8 and 14.

7.3.2.2 Cost Comparison

Table 7.4 presents the optimization results of the two alignments considering existing networks, and compares them to the alignment which does not consider the existing network.

Table 7.4 Comparison of Alignments with and without Considering Existing Road Network

	Alignment 1	Alignment 2	Alignment 3
Considering Existing Network?	No	Network 1	Network 2
Earthwork + Bridge + Tunnel	17.1	13.7	2.0
ROW	12.1	4.2	2.1
Cost to Use Existing Network	0	2.3	4.8
Number of Stations	5	4	4
Total Daily Passengers	3,814	3,197	3155
Travel Time (minutes)	10.8	9.3	9.2
Total Length (miles)	5.8	5.6	5.6
Length on Existing Network	0	2.8	4.4
Initial Cost (M)	54.1	43.0	31.6
Operation/User Cost Saving (M)	-36.5	-43.1	-36.4
Total Cost (M)	17.6	-0.1	-4.8

Compared to Alignment 1, both Alignment 2 and Alignment 3 use existing road networks to reduce the earthwork, bridge/tunnel cost, and ROW cost. Alignment 2 has 2.8 miles along the existing network, saves 3.4M (19.9%) in earthwork and bridge/tunnel cost, and 7.9M (65.3%) in ROW cost. Alignment 3 has 4.4 miles along the existing network, saves 15.1M (88.3%) and 10.0M (82.6%) in earthwork/bridge/tunnel cost and ROW cost, respectively.

Utilizing the existing network decreases the selection opportunity for stations away from the existing network. Both alignments considering an existing network have fewer stations compared to alignment 1, resulting a reduction in total daily passengers. The track lengths and travel times for Alignments 2 and 3 are slightly below those of Alignment 1.

Alignment 2 reduces the total initial cost by 11.1M (20.5%) and Alignment 3 deduces it even more by 22.5M (41.6%). Compared to a total cost of 17.6M for Alignment 1, Alignment 2 and Alignment 3 have total costs of -0.1M and -4.8M, respectively.

7.4 Summary

This chapter presents a practical methodology to take into account existing network in concurrent optimization of railway station locations and alignment. The methodology first defines the criteria for selecting existing links that can accommodate the proposed rail transit alignment, and the costs associated with existing network links and connectors. Then a shortest path algorithm is applied to generate candidate existing paths for any pair of candidate links. A special GA operator is designed to allow the alignment to use an existing network path in order to reduce the total cost in the comprehensive cost estimation framework.

The proposed methodology recognizes the advantages of utilizing existing network to reduce the earthwork cost, bridge/tunnel construction cost, and right-of-way cost. On the other hand, there are also some disadvantages such as the speed

reduction at sharp turns, interference with roadway traffic, and avoidance of stations which are not near the existing network. The model can automatically find a good combination of existing network and new alignment segments to construct an alignment with the minimal total cost. The case study demonstrates the ability of the proposed model to effectively take advantage of the existing network while concurrently optimizing station locations and rail transit alignment.

Chapter 8: Research Summary and Future Work

8.1. Research Summary and Contributions

This thesis focuses on the location and design of the main components of urban rail transit lines, i.e., the location of rail transit stations and the alignment of railway tracks between stations. Based on the needs and constraints of real-world applications, this thesis develops a methodology for concurrently optimizing station locations and track alignments. The modeling framework resolves the essential trade-off between an economically efficient system with low initial and operation cost and an effective system that provides convenient service for the public. The model formulations account for various geometry requirements and real-world design constraints for track alignment and stations plans. The key features of the proposed methodology are presented in Chapter 1.

Chapter 2 provides a comprehensive review of existing research efforts in three aspects of urban rail transit designs, i.e., determination of track alignment, selection of transit stations and evaluation of design alternatives. The review identifies some overlooked critical constraints and unrealistic assumptions in the existing literature, and also reveals the lack of a systemic evaluation framework for accommodating the complex interactions between alignment design and station location.

In response to the identified needs, Chapter 3 proposes a practical rail transit alignment optimization methodology, which can generate alignments that pass through preset station locations while meeting the special geometry constraints at these stations. This base model not only is applicable for rail transit alignment design practice with preset station locations, but also becomes a crucial step towards the development of an integrated optimization model that can concurrently select transit station locations and optimize track alignment between stations. The chapter also proposes a heuristic based on a Genetic Algorithm to efficiently search for solutions that minimize the overall construction cost. The applicability of the base model is extensively examined with a real-world case study and detailed sensitivity analysis in Chapter 4. The numerical results reveal that, even in regions with complex topographical features, the proposed methodology can generate very good alignments that precisely connect the major stations tangentially and satisfy all the geometric constraints.

Chapter 5 presents an extension of the base model, which explicitly incorporates vehicle dynamics in the track alignment design. This extension can realistically simulate the movements of trains and thus can reliably estimate travel time and energy consumption, which are two of the most critical parameters in calculating operation costs and user costs. The generated alignments satisfy geometric constraints both along the alignment and at stations, and achieve the desirable trade-off between the initial cost and the operation/user cost incurred throughout the system's life cycle.

Chapter 6 presents the second extension of the base model, which aims to concurrently optimize station locations and the rail transit alignment connecting these stations. The methodology first constructs the candidate pool of potential rail transit stations based on a comprehensive consideration of various requirements on topological features, accessibility to the existing roadway network, and land availability. The proposed concurrent optimization model, which incorporates a logit model for transit demand forecasting, then selects from the candidate stations and generates alignments between stations to minimize the total cost of the system. The case study demonstrates the advantage of such concurrent optimization over the traditional two-stage optimization, which first locates stations to maximize ridership and then determines the alignment to minimize the construction cost.

Chapter 7 presents the third extension, which aims to take into account the existing network in concurrently optimizing station locations and rail transit alignments. The methodology defines the criteria for selecting existing roadway links that can accommodate the proposed rail transit alignment, and the costs associated with existing network links and connectors. A special GA operator is designed to allow the alignment to use an existing network path in order to reduce the total cost in the comprehensive evaluation framework. The model can automatically find a good combination of existing network links and new alignment segments.

In summary, this research has made the following key contributions:

- This research bridges the present gap between empirical work and academic studies for design and evaluation of urban rail transit systems. It can automatically generate detailed design alternatives in a very efficient manner, while considering complex local features, practical design constrains, and realistic engineering assumptions.

- Developed a concurrent optimization model that incorporates interactions between station location selection and track alignment design, which are usually treated separately in the existing literature. This concurrent optimization methodology addressed the essential trade-off between cost minimization and ridership maximization in the evaluation of design alternatives.

- Accounted for critical geometric design constraints for station layout and track alignment in the station location/alignment optimization process. The generated alignments are consistent in form with real-world rail transit design practice.

- Incorporated vehicle dynamics in track alignment design to realistically simulate the movements of train. The model accounts for the impact of both the horizontal/vertical alignment and the station spacing on operation speed. This proposed methodology can reliably estimate travel time and energy consumption to calculate system operation cost and user cost.

- Developed an optimization model considering existing network in rail transit alignment design. The model can help planners identify the feasible

links of the existing road network that can accommodate the transit alignment, choose the best set among these links, and connect the selected links as well as the transit stations with viable alignments that meet various geometric, topological, environmental and financial constraints.

- Developed a comprehensive cost evaluation framework for comparing design alternatives, which accounts for the tradeoffs between various costs associated with system construction/ operations, and the potential cost savings of introducing the rail transit line.

- Improved the algorithm by incorporating cutting plane spacing as decision variables in the optimization process. This allows unequal spacing between neighboring cutting planes, introduces more flexibility in the track alignments generated, and thus yields design alternatives with lower cost.

8.2. Future Work

Below are some research issues worth considering in future enhancements of the proposed models.

- *Relaxing constraints and repairing infeasible alignments*

Geometry constraint requirements are much stricter for rail track design than for highway design. The higher the design speed is, the more difficult it is to meet all

horizontal alignment and vertical alignment design constraints, such as minimal length of tangent, maximum curvature, minimal spiral curve length, superelevation requirements, and grade constraints. Combined with other environmental and topography constraints, it is possible that very few feasible alignments can be found. A future model should have the ability to repair infeasible alignments by relaxing some geometric constraints, such as reducing design speed at some difficult turns, lowering the requirement of minimal tangent length or curve radius, or allowing reverse curves in extreme conditions. Such changes should be reported to the model-using planner for approval. The model should define the penalty for applying the constraint relaxation and optimize the alignment to reduce the overall system cost accounting for the penalties added.

- *Additional cost components*

The proposed model is an open system with modular construction, which allows the cost evaluation module to be updated or improved in the future. The current cost function includes earthwork cost, bridge/tunnel cost, TOW cost, track cost, train vehicle, station cost, energy cost, operation and maintenance cost, and user cost. The following cost and benefit items can be added to the future model: congestion relief, safety improvement, economic development benefits, and environmental costs. Some additional analytic processes will be needed to represent those costs which are difficult to quantify as monetary values.

- *Improving the algorithm performance*

The current model implements a standard genetic algorithm (sGA). The sGA can be extended to distributed genetic algorithm (dGA) operated on parallel processors. dGA is expected to accelerate the optimization searching process and reduce the chances of getting trapped in local optima.

- *Urban Rail Transit Network Optimization*

The proposed models optimize the station locations and a single track alignment connecting two terminals. The model can be extended to solve rail transit network optimization problems. Instead of using simplified transit lines to connect stations as in most of current models in the literature, the future model should be able to generate realistic rail alignments, account for interactions between different rail transit lines at transfer stations, and reliably estimate network travel time for demand forecasting.

Bibliography

Amtrak 2003, *Track Design Specifications*, Amtrak

AREMA. 2004, *Practical Guide to Railway Engineering*, AREMA

Athanassoulis, G. C. and Calogero, V. 1973, "Optimal Location of a New Highway from A to B: a Computer Technique for Route Planning", *Proceedings of PTRC Seminar on Cost Models and Optimization in Highways*, London

Balaker, T., and Kim C.J. 2006, "Do Economists Reach a Conclusion On Rail Transit," *Econ Journal Watch*, v.3, n.3, pp 551-602

Baltimore Metropolitan Council, "*Baltimore Region Travel Demand Model for Base Year 2000 Task Report 04-01*", January 2004

Baum-Snow, N., and Kahn. M.E. 2005, "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities from 1970 to 2000," *Brookings-Wharton Papers on Urban Affairs*, v.6, pp.147-206

Black, A. 1995, *Urban Mass Transportation Planning*, Mcgraw-Hill Inc, US

Bowman, L. A., and Turnquist M. A. 1981, "Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops," *Transportation Research Part A,* v.15, pp.465-471

Boyle, D. 2006, "Fixed-Route Transit Ridership Forecasting and Service Planning Methods," TCRP Report 66, Transportation Research Board, Washington, D.C.

Bruno, G., Ghiani, G., and Improta, G. 1998, "A Multi-modal Approach to the Location of a Rapid Transit Line," *European Journal of Operational Research*, v.104, pp.321-332

Bruno, G., Gendreau, M., Laporte, G. 2002, "A Heuristic for the Location of a Rapid Transit Line," Computers & Operations Research, v.29, n.1, pp.1-12

Cervero, R., and Zupan, J. 1996, "Commuter and Light Rail Transit Corridors: the Land Use Connection," TCRP Report 16, v.1, Transportation Research Board, Washington, D.C.

Chandler, C., and Hoel, L.A. 2004, "Effects of Light Rail Transit on Traffic Congestion," Research Report, Center for Transportation Studies, University of Virginia

Chatterjee, A., Seaver, W.L., Kiattikomol, V., and Wegmann, F.J. 2002, "Models for Forecasting Urban Transit Ridership for Statewide Analysis," *Compendium of Papers of the 81st TRB Annual Meeting*, Washington, D.C.

Cheng, J.F., and Lee, Y. 2006, "Model for Three-Dimensional Highway Alignment," *Journal of Transportation Engineering*, v.132, n.12, pp.913-920

Chew, E. P., Goh, C. J., and Fwa, T.F. 1989, "Simultaneous Optimization of Horizontal and Vertical Alignments for Highways", *Transportation Research Part B*, v.23, n.5, pp.315-329

Chien, S., and Schonfeld, P. 1997, "Optimization of Grid Transit System in Heterogeneous Urban Environments," *Journal of Transportation Engineering*, v.123, n.1, pp.28-35

Chu, X. 2004, "Ridership Models at the Stop Level," Research Report, University of South Florida

Cline, J.C. 1986, "Delay Impacts of Light Rail Transit Grade Crossings," M.S. Thesis, Texas A&M University

Cockerill, L. and Stanley, D. 2002, "How Will the Centerline Affect Property Values in Orange County: A Review of the Literature and Methodological Approaches for Future Consideration," Research Report, California State University

Compin, N.S. 1999, "The Four Dimensions of Rail Transit Performance: How Administration, Finance, Demographics and Politics Affect Outcomes," Dissertation, University of California at Irvine

Conger, M. 2007, "Recent and Planned Improvements to the Knoxville Regional Travel Demand Forecasting Model," *Presented at the TMIP Web Seminar on Travel Forecasting Modeling Activities at MPOs*

Current, J.R., ReVelle, C.S., and Cohon, J. 1985, "The Maximum Covering/Shortest Path Problems: A Multiobjective Network Design and Routing Formulation," *European Journal of Operational Research*, v.21, pp.189-199

Current, J., ReVelle, C., Cohon, J. 1987, "The Median Shortest Path Problem: A Multiobjective Approach to Analyze Cost vs. Accessibility in the Design of a Transportation Network," *Transportation Science*, v.21, n.3, pp.188-197

Davies, J.V. 2007, "Construction of A Rapid Transit Railroad in Relation to the Handling of Passengers as Illustrated by the Hudson And Manhattan Railroad," *Proceedings of the Engineers' Club of Philadelphia*, pp.316-336

de Smith, M.J. 2006, "Determination of Gradient and Curvature Constrained Optimal Paths," *Computer-Aided Civil and Infrastructure Engineering*, v.21, n.1, pp.24-38

Dufourd, H., Gendreau, M., And Laporte, G. 1996, "Locating a Transit Line Using Tabu Search," *Location .Science*, v.4, n.1/2, pp.l-19

Easa, S.M. 1988, "Selection of Roadway Grades that Minimize Earthwork Cost Using Linear Programming", *Transportation Research Part A*, v.22, n.2, pp.121-136

Eppli, M., and Tu, C.C. 2000, "Valuing the New Urbanism: the Impact of New Urbanism on Prices of Single-Family Homes," Research Report, Urban Land Institute, Washington D.C.

Faghri, A., Lang, A, Hamad, K., and Henck, H. 2002, "Integrated Knowledge-Based Geographic Information System for Determining Optimal Location of Park-and-Ride Facilities," *Journal of Urban Planning and Development*, v.128, n.1, pp.18-41

Farhan, B., and Murray, A.T. 2003, "A GIS-Based Approach for Delineating Market Areas for Park-and-Ride Facilities," *Compendium of Papers of the 82nd TRB Annual Meeting,* Washington, D.C.

Fielding, G., and Glauthier, R. 1976, "Distribution and Allocation of Transit Subsidies in California," Working Paper, University of California at Irvine

Flyvbjerg, B., Holm, M.K.S., and Buhl, S.L. 2005, "How (In)accurate are Demand Forecasts in Public Works Projects: the Case of Transportation," *Journal of the American Planning Association*, v.71, n.2, pp.131-146

Frank, L., Kavage S., and Litman T. 2006, "Promoting Public Health through Smart Growth: Building Healthier Communities through Transportation and Land Use Policies," Research Report, Victoria Transport Policy Institute, Canada

FTA 2002, "Transit Performance Monitoring System (TPMS) Results", Accessed via Internet on March 1, 2008 at http://www.fta.dot.gov/publications/reports/ publications_5677.html

Fwa, T. F. 1989, "Highway Vertical Alignment Analysis by Dynamic Programming", *Transportation Research Record*, n.1239, pp.1-9

Fwa, T. F., Chan, W. T., and Sim, Y. P. 2002, "Optimal Vertical Alignment Analysis for Highway Design," *Journal of Transportation Engineering*, v.128, n.5, pp.395-402

FWTA 2007, *Southwest-to-Northeast Rail Corridor Station Planning Process and Guidelines*, FWTA

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

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

Goh C. J., Chew, E. P., and Fwa, T. F. 1988, "Discrete and Continuous Models for Computation of Optimal Vertical Highway Alignment", *Transportation Research Part B*, v.22, n.6, pp.399-409

Google Earth, "*http://www.panoramio.com/photo/54305354*", June 16, 2011

Hartgen, D., and Horner, M. W. 1997, "A Route-level Transit Ridership Forecasting Model for Lane Transit District: Eugene, Oregon," Research Report Center for Interdisciplinary Transportation Studies, Charlotte, NC

Hassan, Y., Easa, S.M., Abd El Hlim, A.O. 1997, "Design Considerations for Combined Highway Alignment," Journal of Transportation Engineering, v.123, n.1, pp.60-68

Haughwout, A.F. 2000, "The Paradox of Infrastructure Investment," *Brookings Review*, v.18, n.3, pp.40-43

Hay, W.W. 1982, *Railroad Engineering*, Wiley, John & Sons

Hayman, R, W. 1970, "Optimization of Vertical Alignment for Highways through Mathematical Programming", *Highway Research Record*, n.306, pp.1-9

Hess, D., and Ong, P. 2001, "Traditional Neighborhoods and Auto Ownership," Working Paper, University of California at Los Angeles

Hilton, G. 1976, "What Did We Give Up with the Big Red Cars? Transportation Alternatives for Southern California," Working paper, University of California at Los Angeles

Hogan, J. D. 1973, "Experience with OPTLOC: Optimum Location of Highway by Computer", *Proceedings of PTRC Seminar on Cost Models and Optimization in Highways*, London

Horner, M.W., and Grubesic, T.W. 2001, "A GIS-based Planning Approach to Locating Urban Rail Terminals," *Transportation*, v.28, pp.55-77

Howard, B. E., Bramnick, Z., and Shaw, J. F. B. 1968, "Optimum Curvature Principle in Highway Routing", *Journal of the Highway Division,* v.94, n.1, pp.61-82

I-70 Coalition. 2008, "Technical Specifications of a Transit Center: Land Use Planning Study for Rail Transit Alignment through the I-70 Corridor," Project Report accessible at http://rockymountainrail.org/documents/I70_Coalition_Task_ 9_ Tech_ Memo_120508.pdf

ITE 1992, *Guidelines for Design of Light Rail Grade Crossings*, Institute of Transportation Engineers

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

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

Jha, M. K., Schonfeld, P., Jong, J.C., and Kim, E. 2006, *Intelligent Road Design*, WIT Press: Southhampton, UK

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

Jha, M.K., Karri, G.A.K., and Kuhn, W. 2011, "New Three-Dimensional Highway Design Methodology for Sight Distance Measurement," *Transportation Research Record, Journal of the Transportation Research Board,* v.2262, pp.74-82

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

Jong J.C., and Schonfeld, P. 2003, "An Evolutionary Model for Simultaneously Optimizing Three-dimensional Highway Alignments," *Transportation Research Part B*, v.37, pp.107–128

Kang, M.W., Schonfeld, P., and Jong J.C. 2007, "Highway Alignment Optimization through Feasible Gates," *Journal of Advanced Transportation*, v.41, n.2, pp.115-144

Kang, M.W. 2008, "An Alignment Optimization Model for A Simple Highway Network," Ph.D. Dissertation, University of Maryland at College Park

Kemp, M. A. 1981, "A Simultaneous Equations Analysis of Route Demand, Supply, and Its Application to the San Diego Bus System," Research Report n.1470-2, the Urban Institute, Washington D.C.

Kenworthy J. and Laube, F. 2000, "The Role of Light Rail in Urban Transport Systems: Winning back Cities from the Automobile," *Presented at the 5th Light Rail Conference*, Melbourne, Australia

Kikuchi S., and Miljkovic, D. 2001, "Use of Fuzzy Inference for Modeling Prediction of Transit Ridership at Individual Stops," *Transportation Research Record*, n.1774, pp.25-35

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

Kim, D.N., and Schonfeld, P. 1997, "Benefits of Dipped Vertical Alignments for Rail Transit Routes," *Journal of Transportation Engineering*, v.123, n.1, pp.20-27

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

Kim, E., Jha, M.K., and Schonfeld, P. 2004, "Intersection Construction Cost Functions for Alignment Optimization," *Journal of Transportation Engineering*, v.130, n.2, pp.194-203

Kim, K. and Chien, S.I. 2010, "Simulation-Based Analysis of Train Controls under Various Track Alignments," Journal of Transportation Engineering, v.136, n.11, pp.937-948

Kimpel T.J., Strathman J.G., and Dueker, K.J. 2000, "Time Point-Level Analysis of Passenger Demand and Transit Service Reliability," Working paper, Portland State University

Kimpel T.J. 2001, "Time Point-Level Analysis of Transit Service Reliability and Passenger Demand," Dissertation, Portland State University

Klauder, L.T., Chrismer, S.M., and Elkins, J. 2002, "Improved Spiral Geometry for High-Speed Rail and Predicted Vehicle Response," *Transportation Research Record,* n.1785, pp.41-49

Kufver, B. 1997, "Optimization of Single Horizontal Curves in Railway Alignments," TRITA-FKT Report 1997:45, Swedish National Road and Transport Research Institute, Sweden

Kufver, B. 1998, "High Speed Railways: Alignment Optimization with Vehicle Reactions Taken into Consideration," *Nordic Road & Transport Research*, n.2, pp.4-6

Kufver, B. and Andersson, E. 1998, "Optimisation of Lengths of Transition Curves with Vehicle Reactions Taken into Consideration," *Proceedings of Comprail 1998: Advances in Transport Series*, v.2, pp.33-42

Kuhn, W. and Jha M.K. 2011, "Methodology for Checking Shortcomings in the Three-Dimensional Alignment," *Transportation Research Record, Journal of the Transportation Research Board,* n.2262, pp.13-21

Lai, X. 2002, "Research on GIS Based Urban Rail Transit Network Planning and Passenger Forecasting System," Master Thesis, Tsinghua University, China

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

Lai, X. and Schonfeld, P. 2012 (In Press), "Optimization of Rail Transit Alignments Considering Vehicle Dynamics," *Transportation Research Record, Journal of the Transportation Research Board*

Lane, C., DiCarlantonio, M., and Usvyat, L. 2006, "Sketch Models to Forecast Commuter and Light Rail Ridership: Update to TCRP Report 16," *Compendium of Papers of the 85th TRB Annual Meeting,* Washington, D.C.

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

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

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

Lave, C. 1998, "25 Years of U.S. Energy Policy Successes, Failures, and Some General Lessons for Public Policy," *Transportation Quarterly*, v.52, n.4, pp.7-13

Lee, Y., and Cheng, J. F. 2001(a), "A model for Calculating Optimal Vertical Alignments of Interchanges," *Transportation. Research Part B*, v.35, n.5, pp.423-445

Lee, Y., and Cheng, J. F. 2001(b), "Optimizing Highway Grades to Minimize Cost and Maintain Traffic Speed," *Journal of Transportation. Engineering*, v.127, n.4, pp.303–310

Lee, Y.-J. and Vuchic, V.R. 2005, "Transit Network Design with Variable Demand," *Journal of Transportation Engineering,* v.131, n.1, pp.1-10

Lewis, D., and Williams, F. L. 1999, *Policy and Planning as Public Choice: Mass Transit in the United States*, Ashgate Publishing Ltd

Lewis, D., and Bekka, K. 2000, "Transit Benefits 2000 Working Papers: A Public Choice Policy Analysis." FTA Policy Paper, Federal Transit Administration, U.S. Department of Transportation, Washington, D.C.

Litman, T. 2003, "Evaluating Criticism of Smart Growth," Research Report, Victoria Transport Policy Institute, Canada

Litman, T. 2006(a), "Rail Transit in America: A Comprehensive Evaluation of Benefits," Research Report, Victoria Transport Policy Institute, Canada

Litman, T. 2006(b), "Comprehensive Evaluation of Rail Transit Benefits: a Comment," Working paper, Victoria Transport Policy Institute, Canada

Marin, A., and Garcia-Rodenas, R. 2009, "Location of Infrastructure in Urban Railway Networks," *Computers & Operations Research*, v.36, n.5, pp.1461-1477

Marshall, N., and Grady, B. 2006, "Sketch Transit Modeling Based on 2000 Census Data," Transportation Research Record, n.1986, pp.182-189

MARTA 2007, "Detailed Screening Results and Selection of Locally Preferred Alternative for Beltline Alternatives Analysis," Project Report, Metropolitan Atlanta Rapid Transit Authority

MTA 2007, "Purple Line – Bethesda to Silver Spring Segment: System Design Criteria,", Maryland Transit Administration

McBrayer D.B. 2003, "Light Rail Transit and Other Modes: Blurring the Light Rail Transit–Bus Rapid Transit Boundaries," *Transportation Research Circular E-C058: 9th National Light Rail Transit Conference*, pp135-147

Meyer, M.., and Miller, E., 2001, *Urban Transportation Planning: A Decision-Oriented Approach*. McGraw-Hill: New York

Miller, J., Robison, H., and Lahr, M. 1999, "Estimating Important Transportation-Related Regional Economic Relationships in Bexar County, Texas," Research Report, Victoria Transport Policy Institute, Canada

Mogridge, M. 1990, *Travel in Towns: Jam Yesterday, Jam today, and Jam Tomorrow?* The Macmillan Press, Ltd.: London, England

Moreb, A. A. 1996, "Linear Programming Model for Finding Optimal Roadway Grades that Minimize Earthwork Cost," *European Journal of Operational Research*, v.93, n.1, pp.148–154

Murchland, J. D. 1973, "Methods of Vertical Profile Optimisation for an Improvement to an Existing Road", *Proceedings of PTRC Seminar on Cost Models and Optimization in Highways*, London

NCTCOG. 2007, "Dallas-Fort Worth Regional Travel Model (DFWRTM): Model Description," Project Report, North Central Texas Council of Governments

Nelson, P., Baglino, A., Harrington, W., Safirova, E., and Lipman, A. 2007, "Transit in Washington, D.C.: Current Benefits and Optimal Level of Provision," Journal of Urban Economics, v.62, n.2, pp.231-251

Nicholson, A. J., Elms, D.G., and Williman, A. 1976, "A Variational Approach to Optimal Route Location", *Highway Engineers*, v.23, pp.22-25

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

Peng, Z.R., Dueker, K.J., Strathman, J., and Hopper, J. 1997, "A Simultaneous Route-Level Transit Patronage Model: Demand, Supply, and Inter-Route Relationship," *Transportation*, v.24, n.2, pp.159–181

Podobnik, B. 2002, "The Social and Environmental Achievements of New Urbanism: Evidence from Orenco Station," Working Paper, Lewis and Clark College, Portland, Oregon

Pushkarev B., Zupan, J.M., and Cumella, R.S. 1982, *Urban Rail in America: An Exploration of Criteria for Fixed-Guideway Transit*, Indiana University Press: Bloomington

Puy Huarte, J. 1973, "OPYGAR: Optimisation and Automatic Design of Highway Profiles", *Proceedings of PTRC Seminar on Cost Models and Optimization in Highways*, London

Renne, J.L. 2005, "Transit-Oriented Development: Measuring Benefits, Analyzing Trends, And Evaluating Policy," Dissertation, Rutgers State University

ReVelle, C. S., Whitlatch, E. E., and Wright, J. R. 1997, *Civil and Environmental Systems Engineering*, Prentice Hall, New Jersey

Richmond, J. 2001, "A Whole-System Approach to Evaluating Urban Transit Investments," *Transport Reviews*, v.21, n.2, pp.141-180

Rousseau, G. 2007, "From Trip-Based To Tour-Based: A Systematic & Incremental Way of Migrating to Activity-Based & Tour-Based Modeling - The Atlanta Regional Commission Experience," *Presented at the TMIP Web Seminar on Travel Forecasting Modeling Activities at MPOs*

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

Samanta, S. 2008, "Models and Algorithms for a Rail Transit Line Alignment Using GIS and Genetic Algorithm" Ph.D. Dissertation, Morgan State University

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

Samanta, S. and Jha, M.K 2011, "Modeling a Rail Transit Alignment Considering Different Objectives," *Transportation Research, Part A*, v.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 Systems*, v.3, n.1, pp.13-36

Sanchez, E. 1994, "Three-Dimensional Analysis of Sight Distance on Interchange Connectors," *Transportation Research Record*, n.1445, pp.101-108

Saur, G.J., Lee, R., and Gray, C. 2004, "New Method for Transit Ridership Forecasting," *Presented at the 2004 Annual Meeting of the Institute of Transportation Engineers*

Shapiro, R.J., Hassett, K.A., and Arnold, F.S. 2002, "Conserving Energy and Preserving the Environment: The Role of Public Transportation," Research Report, American Public Transportation Association

Shaw, J. F. B. and Howard, B. E. 1981, "Comparison of Two Integration Methods in Transportation Routing", *Transportation Research Record*, n.806, pp.8-13

Shaw, J. F. B. and Howard, B. E. 1982, "Expressway Route Optimization by OCP", *Journal of Transportation Engineering*, v.108, n.3, pp.227-243

Smith, B.L., and Lamm, R. 1994, "Coordination of Horizontal and Vertical Alignment with Regard to Highway Esthetics," *Transportation Research Record*, n.1445, pp.73-85

Smith, J.J., and Gihring, T.A. 2006, "Financing Transit Systems through Value Capture: an Annotated Bibliography," *American Journal of Economics and Sociology*, v.65, n.3, pp.751-786

Stopher, P.S. 1992, "Development of a Route-Level Patronage Forecasting Method," *Transportation*, v.19, n.3, pp. 201–220

Switzer, C. 2002, "The Center Commons Transit Oriented Development: A Case Study," Master Thesis, Portland State University

Taylor, B.D., and Fink, C.N.Y. 2003, "The Factors Influencing Transit Ridership: a Review and Analysis of the Ridership Literature," Working Paper, University of California at Los Angeles

Taylor, B.D., Miller, D., Iseki, H., and Fink C. 2004, "Analyzing the Determinants of Transit Ridership Using a Two-Stage Least Square Regression on a National Smaple of Urban Areas," *Compendium of Papers of the 83rd TRB Annual Meeting*, Washington, D.C.

Thompson, G.L. 1997, "Achieving Suburban Transit Potential: Sacramento Revisited," *Transportation Research Record*, n.1571, pp.151-161

Thompson, G.L., and Brown, J.R. 2006, "Explaining Variation in Transit Ridership in U.S. Metropolitan Areas Between 1990 and 2000: Multivariate Analysis," *Transportation Research Record*, n.1986, pp.172-181

TRB 2000, "Track Design Handbook for Light Rail Transit," TCRP Report 57, Transportation Research Board, Washington, D.C.

TRB 2002, "Estimating the Benefits and Costs of Public Transit Projects: A Guidebook for Practitioners," TCRP Report 78, Transportation Research Board, Washington, D.C.

Trietsch, D. 1987a, "A Family of Methods for Preliminary Highway Alignment", *Transportation Science*, v.21, n.1, pp.17-25

Trietsch, D. 1987b, "Comprehensive Design of Highway Networks", *Transportation Science*, v.21, n.1, pp.26-35

Turner, A. K. and Miles, R. D. 1971, "A Computer-Assisted Method of Regional Route Location", *Highway Research Record*, n.348, pp.1-15

Turner, A.K. 1978, "A Decade of Experience in Computer Aided Route Selection," *Photogrammetric Engineering and Remote Sensing*, v.44, pp.1561-1576

Vuchic, V.R., and Newell, G.F. 1968, "Rapid Transit Interstation Spacings for Minimum Travel Time," *Transportation Science*, v.2, n.4, pp.303-339

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

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

Vuchic, V.R. 1999, *Transportation For Livable Cities*, CUPR Press

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

Walters, G., and Cervero, R. 2003, "Forecasting Transit Demand in a Fast Growing Corridor: The Direct-Ridership Model Approach," Working paper, University of California at Berkeley

Wang, J.Y.T., Yang, H., and Lindsey, R. 2004, "Locating and Pricing Park-and-Ride Facilities in a Linear Monocentric City with Deterministic Mode Choice," *Transportation Research Part B*, v.38, n.8, pp.709-731

Weinstein, B.L., and Clower, T.L. 2003, "An Assessment of the DART LRT on Taxable Property Valuations and Transit Oriented Development," Research Report, University of North Texas

Wey, W., and Chang, Y. 2007, "A Comparative Location Study for a Joint Development Station of MRT in Taichung City of Taiwan," *Compendium of Papers of the 86th TRB Annual Meeting,* Washington, D.C.

Wirasinghe, S.C., and Ghoneim, N.S. 1981, "Spacing of Bus-Stops for Many to Many Travel Demand," *Transportation Science*, v.15, n.3, pp.210-221

WMATA 1976, "Rapid Transit System: Manual of Design Criteria," Washington Metropolitan Area Transit

Wohl, M., and Hendrickson, C. 1985, *Transportation Inverstment and Pricing Principles*, John Wiley & Sons Inc

Yeh, S.T. 2003, " Integrated Analysis of Vertical Alignments and Speed Profiles for Rail Transit Routes," Master Thesis, University of Maryland at College Park

Zaretsky, Adam M. 1994, "Riding the Rails: A Look at Light Rail Transit," *The Regional Economist*, October 1994, pp.4-9

Zhao F., Chow, L., Li, M., and Liu X. 2005, "A Transit Ridership Model Based on Geographically Weighted Regression and Service Quality Variables," Research Report, Florida International University