

Concurrent Optimization of Rail Transit Alignments and Station Locations

Xiaorong Lai¹  · Paul Schonfeld²

Received: 9 November 2015/Revised: 4 February 2016/Accepted: 14 February 2016/Published online: 16 March 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Designing urban rail transit systems is a complex problem, which involves the determination of station locations, track geometry, and various other system characteristics. Most of the previous rail transit route optimization studies have focused on the alignment design between predetermined stations, whereas a practical design process has to account for the complex interactions among railway alignments and station locations. This paper proposes a methodology for concurrently optimizing station locations and the rail transit alignment connecting those stations, by accommodating multiple system objectives, satisfying various design constraints, and integrating the analysis models with a geographical information system database. The methodology incorporates demand and station costs in the evaluation framework and employs a genetic algorithm for optimizing the decision variables for station locations, station types, and track alignments. It is expected that transit planners may greatly benefit from the proposed methodology, with which they can conveniently and efficiently optimize candidate alternatives. The Baltimore Red Line is used as a case study to demonstrate how the model can find very good solutions in regions with complex geography.

Keywords Rail transit · Concurrent optimization · Station location · Track alignment

1 Introduction

As a high-capacity transportation mode that is free from surface road congestion, urban rail transit has witnessed significant expansion in many cities during recent decades despite the substantial investment needed for its construction and maintenance. Designing urban rail transit systems is a complex problem, which involves the determination of station locations, track geometry, and various other system characteristics. Currently, the design of urban rail transit systems is mostly approached empirically with a trial-and-error process: planners develop alternatives subject to design specifications and local conditions, and then evaluate these alternatives based on project budget as well as system performance criteria. Such a design process is time-consuming and cannot guarantee that its results are even close to optimal.

In response, various optimization models have been proposed to establish alignments that meet various geometry requirements. The early work focused especially on modeling techniques, such as calculus of variations [1, 2], numerical search [3, 4], linear programming [5, 6], network optimization [7–9], and dynamic programming [3, 8, 10, 11]. Recent studies advanced the methodology by removing unrealistic assumptions and incorporating real-world constraints. Researchers at the University of Maryland have proposed a series of GA-based alignment optimization models. Jong [12] and Jong and Schonfeld [13] first demonstrated the concept in highway alignment optimization, featuring its comprehensive cost function and its consistency with engineering practice in the generated

✉ Xiaorong Lai
xlai@rkk.com

Paul Schonfeld
pschon@umd.edu

¹ Rummel, Klepper & Kahl, LLP, 81 Mosher Street, Baltimore, MD 21217, USA

² Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, USA

Editor: Xuesong Zhou

alignments. Jha [14], Jha and Schonfeld [15] and Jha et al. [16] extended the model by integrating a GIS to better accommodate the complex topological and environmental features. Kim [17] and Kim et al. [18] developed methods for incorporating the cost of major structures, i.e., bridges and tunnels. Kang et al. [19] further improved the GA-based solution algorithm by introducing the Feasible Gates approach.

Jha et al. [20] extended the previous highway models to railway alignment optimization. Lai and Schonfeld [21] presented a practical rail transit alignment optimization method to account for vehicle dynamics, which aims to balance the initial construction cost with the operation and user costs recurring throughout the system's life cycle. Kim et al. [22, 23] focused on vertical alignments between rail transit stations that exploited gravity to help accelerate and decelerate trains. Most of these publications on optimizing urban rail transit designs focused on the alignment optimization between two or more predetermined stations, whereas the selection of station locations may actually be even more challenging. 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.

This paper proposes a methodology for concurrently optimizing station locations and the rail transit alignment connecting these stations, by accommodating multiple system objectives, formulating various design constraints, and integrating the analysis models with a geographical information system (GIS) database. The methodology incorporates demand and station costs in the evaluation framework and employs a genetic algorithm (GA) for optimizing the decision variables for station locations, station types, and track alignments. It is expected that transit planners may greatly benefit from the proposed methodology, with which they can conveniently and efficiently optimize alternatives.

2 Methodology

Figure 1 illustrates the framework of the proposed methodology for concurrently optimizing rail transit alignment and station locations, and also for helping planners select the type of each station.

Following the engineering practice, this paper models the 3-dimensional rail track alignment with three separate components: the selection of stations, including station sequence and station type, the horizontal alignment that defines the track's path on the *XY* plane, and the vertical alignment that defines the elevation along the horizontal alignment. The model also incorporates various geometry constraints derived from engineering practice [24], using the

cutting-plane concept [12] to define the Point of Intersection. The simulation of vehicle dynamics is based on the essential train dynamics equations in Hay [25], which divide the train operation between each pair of neighboring stations into acceleration, cruising, and braking stages, and then calculates train energy consumption and travel time along a given track alignment with an iterative process [21].

The remaining sections will detail three other key steps, i.e., the generation of the candidate station pool, the mode choice model, and the calculation of the cost function.

2.1 Generation of the Candidate Pool of Potential Rail Transit Stations

The proposed concurrent optimization model directly addresses the tradeoffs between ridership and cost in selecting station locations and generating the alignment in between. As it is impractical to check every point in the study area as potential station sites, this section presents 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 concurrent station and alignment optimization model to identify the best station locations.

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

1. Stations cannot be located within infeasible areas (e.g., lakes or rivers), environmentally sensitive areas (e.g., wetland or residence of protected species), or historically sensitive areas (e.g., churches or cemeteries).
2. Stations should have the potential to attract considerable ridership, which can be realized in three ways. First, the catchment areas of stations could cover a sufficient number of households or employment locations. However, the size of the catchment area is related to the selected station type: walking-based stations have a shorter radius compared to park-and-ride stations. Secondly, stations may 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 land reserved for future transit-oriented development, may 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

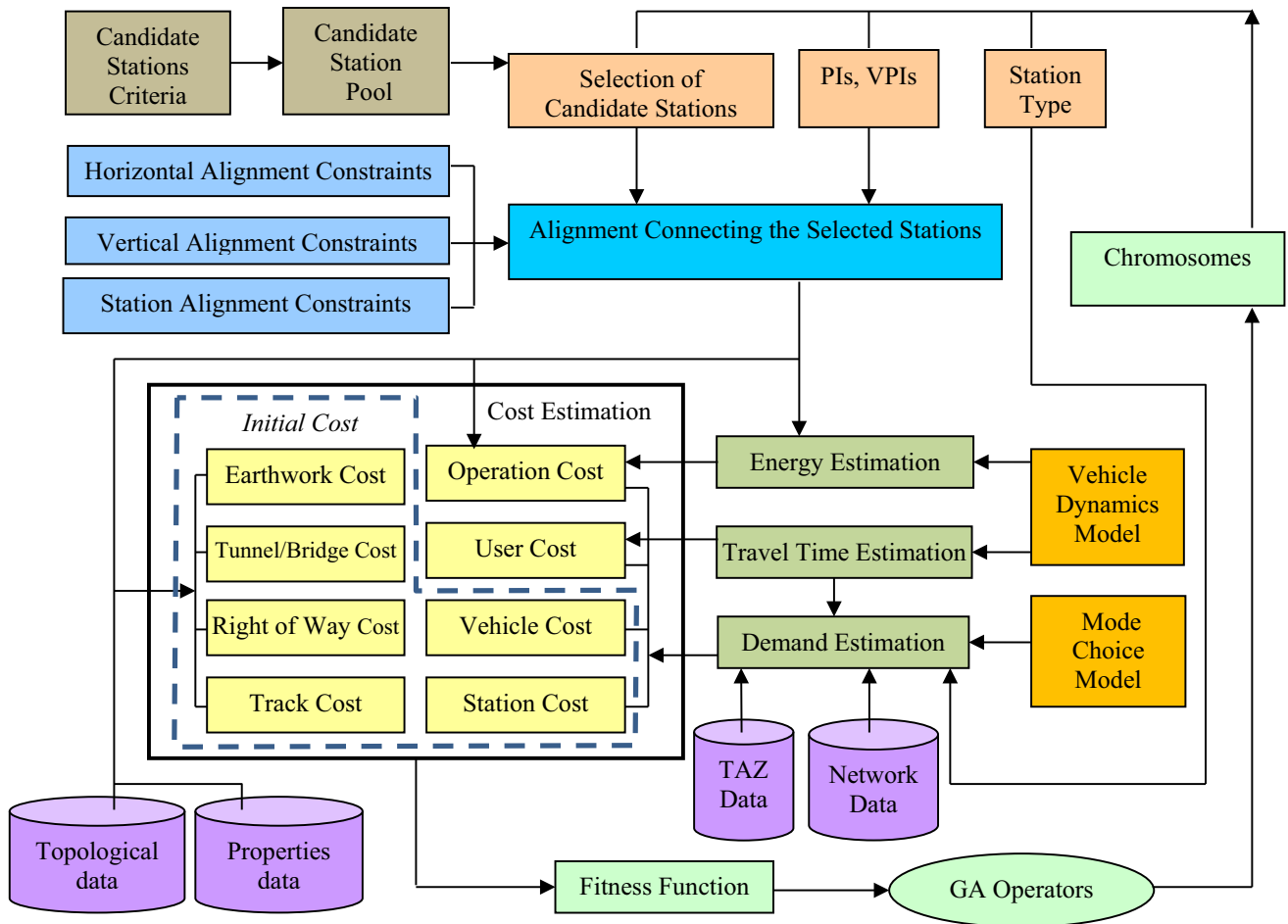


Fig. 1 Framework of the concurrent optimization model

neighborhoods with expensive right-of-way cost. The park-and-ride stations should not be too close to downtown; otherwise commuters would be unlikely to use rail transit for very short trips.

In accordance with these selection criteria, the following procedures are used here 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—infesible; 1—feasible)
 S_p^i , feasibility of grid i to be park-and-ride station (0—infesible; 1—feasible)
 Set $S_w^i = 0$, $S_p^i = 0$
- Step 2 Create a layer Ω_i for infesible areas, which combines wetlands, historic districts, historically sensitive areas, and 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

$$\Omega_f = \Omega_s \cap \bar{\Omega}_i \tag{1}$$

- Step 4 Find all grids with the number of households within walking distance above a threshold

For each $G_i \in \Omega_f$,

$$\text{set } S_w^i = 1 \text{ if } \left(\sum_j H_j |R_j^i| \leq R_w \right) \geq H_w \cap \bar{\Omega}_i, \tag{2}$$

where H_j is the number of households in 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 above a threshold

For each $G_i \in \Omega_f$,

$$\text{set } S_p^i = 1 \text{ if } \left(\sum_j H_j |R_j^i \leq R_p \right) \geq H_p, \quad (3)$$

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 above a threshold

For each $G_i \in \Omega_f$,

$$\text{set } S_w^i = 1 \text{ if } \left(\sum_j E_j |R_j^i \leq R_w \right) \geq E_w \quad (4)$$

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 $D_j^i \leq R_w$ exceeds 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 right-of-way (ROW) cost

For each $G_i \in \Omega_f$,

$$\text{set } S_w^i = 0 \text{ if } \left(\sum_j C_j |L_j^i \leq L_w \right) \geq C_w,$$

$$\text{set } S_p^i = 0 \text{ if } \left(\sum_j C_j |L_j^i \leq L_p \right) \geq C_p, \quad (5)$$

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 above 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 the 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 the 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 , based 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 encodes 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 must satisfy the following constraints:

(a) Minimum number of stations N_L

$$\sum_i X_i \geq N_L \quad (6)$$

(b) Maximum number of stations N_U

$$\sum_i X_i \leq N_U \quad (7)$$

(c) Minimum spacing between any two selected stations

$$\forall X_i = 1, X_j = 1, i \neq j: Y_{ij} \geq Y_{\min}, \quad (8)$$

where Y_{ij} is the distance along the alignment from Station i to Station j , Y_{\min} is the minimum spacing required between stations

(d) Maximum spacing between any two selected stations

$$\forall X_i = 1, X_j = 1, i \neq j: Y_{ij} \leq Y_{\max}, \quad (9)$$

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

- (e) Minimum distance to depart from the starting terminal

$$\forall X_i = 1, X_j = 1, j > i : L_j^0 - L_i^0 \geq L_0, \tag{10}$$

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

- (f) Minimum distance to approach the end terminal

$$\forall X_i = 1, X_j = 1, j > i : L_i^1 - L_j^1 \geq L_1, \tag{11}$$

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

These decision variables and constraints are then incorporated into a GA-based solution heuristic for generating alignments that connect the selected stations.

2.2 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 [26]. Such methods are quite simplified compared to various transit ridership forecasting models that are used in rail transit planning studies. This paper incorporates in the proposed concurrent optimization framework a discrete choice model, which is a widely accepted transit ridership forecasting model in real-world practice.

2.2.1 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. In this paper, it is assumed that the total trips matrix is known from external regional demand forecasting models, and thus

aggregate choice models are employed. 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 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.

2.2.2 Multinomial Logit Choice Model for Pedestrian-Oriented Stations

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

$$U_m = V_m + \varepsilon_m, \tag{12}$$

where V_m is the representation of utility using observed variables, ε_m is the unknown 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 it being chosen. The point at which the increase in representative utility has the greatest effect on the probability of it 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 Fig. 2.

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

$$P_{ijm} = \frac{e^{V_{ijm}}}{\sum_m e^{V_{ijm}}}. \tag{13}$$

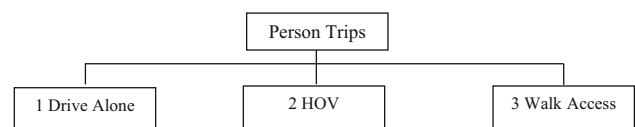


Fig. 2 Multinomial logit choice model for pedestrian-oriented stations

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

$$P_{ijm} = \frac{e^{\beta'x_{ijm}}}{\sum_m e^{\beta'x_{ijm}}} \tag{14}$$

2.2.3 Nested Logit Choice Model for Park-and-Ride Stations

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. An 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

$$U_m = W_k + Y_m + \varepsilon_m \tag{15}$$

$$V_m = W_k + Y_m. \tag{16}$$

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 Fig. 3.

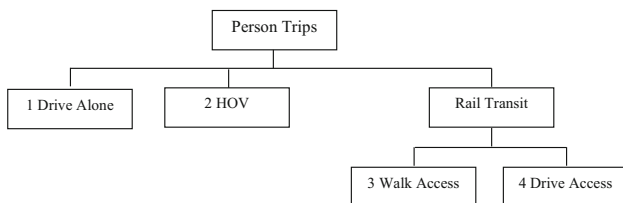


Fig. 3 Nested logit model for park-and-ride stations

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, $P_{ijm|B_k}$, which can be obtained by using MNL model

$$P_{ijm} = P_{ijm|B_k} \times P_{ijB_k} \tag{17}$$

$$P_{ijm|B_k} = \frac{e^{Y_{ijm}/\lambda_k}}{\sum_{l \in B_k} e^{Y_{ijl}/\lambda_k}} \tag{18}$$

$$P_{ijB_k} = \frac{e^{W_k + \lambda_k I_{ijk}}}{\sum_{l=1}^K e^{W_l + \lambda_l I_{ijl}}} \tag{19}$$

$$I_{ijk} = \ln \left(\sum_{m \in B_k} e^{Y_{ijm}/\lambda_k} \right). \tag{20}$$

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 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).

2.3 Estimation of System Cost

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

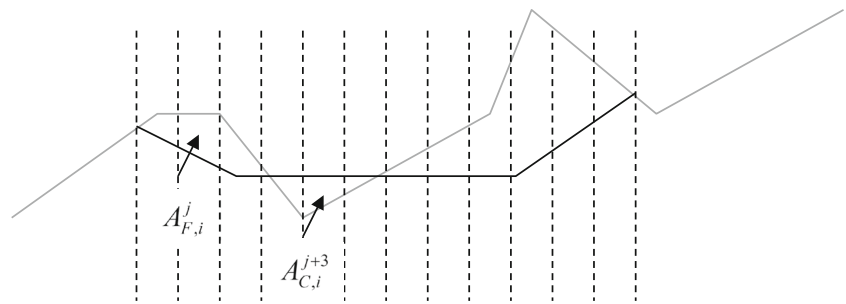
$$C_{net} = C_c - S_o - S_U. \tag{21}$$

2.3.1 Initial Cost

The initial cost of the alignment is the capital cost, which includes earthwork costs C_E , bridge costs C_B , tunnel costs C_T , right-of-way costs C_R , track costs C_L , train vehicle costs C_V , and station costs C_s :

$$C_c = C_E + C_B + C_T + C_R + C_L + C_V + C_s. \tag{22}$$

Fig. 4 Earthwork of a typical cut/fill section



Following engineering practice, this paper applies the typical track cross sections, as shown in Fig. 4, at an equal spacing L_{CS} along the horizontal alignment on the corresponding elevation from the vertical alignment. It then uses the elevation data from the GIS database to characterize the cut/fill sections that contribute earthwork costs and bridge/tunnel sections that contribute to structure costs. The total earthwork cost for a typical cut/fill section is calculated as follows:

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

$$C_E = K_C \sum_i E_{C,i} + K_F \sum_i E_{F,i} + K_l \max(E_N, 0) - k_b \min(E_N, 0) \quad (23)$$

Here E_N is the net earthwork, C_E is the total earthwork cost, s_e is earth shrinkage factor, K_C is unit cutting cost, K_F is unit filling cost, and K_l and k_b are unit transportation cost for, respectively, moving earth to a landfill and from a borrow pit. The cut volume $E_{C,i}$ and fill volume $E_{F,i}$ are calculated for each cross section i by stratifying each section with very small intervals and calculating the cut area $A_{C,i}^j$ and fill area $A_{F,i}^j$ between the proposed and the existing ground for each stratum j .

$$E_{C,i} = 0.5 \times \left(\sum_j A_{C,i}^j + \sum_j A_{C,i+1}^j \right) \times L_{CS}$$

$$E_{F,i} = 0.5 \times \left(\sum_j A_{F,i}^j + \sum_j A_{F,i+1}^j \right) \times L_{CS} \quad (24)$$

The total structure cost includes both bridge cost and tunnel cost. For each bridge i , an enumeration method is used to find the optimal span length L_{Bi} that minimizes the sum of superstructure C_{Bi}^U and substructure costs C_{Bi}^L . 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 depends on the predefined unit cost for tunnel excavation, the area of tunnel cross sections and the tunnel length. To calculate the right-of-way cost, this paper 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 outside the edge of earthwork with a buffer width. For bridges, the edge points are outside the bridge cross section with a buffer width. Tunnel sections require no right-of-way. The track cost depends only on the track length and a unit track installation cost. The vehicle costs C_V are the product of the number of trains N_T needed in the fleet, the number of cars per train N_c , and the cost per car K_V in millions:

$$C_V = N_T \times N_c \times K_V \quad (25)$$

Assuming a fixed headway H in train schedule, the number of needed trains N_T can be calculated as the round trip travel time divided by the headway:

$$N_T = \frac{T_R}{H} \quad (26)$$

The above formulations of train vehicle costs indicate that lower travel times require fewer trains, and thus decrease the cost of purchasing vehicles. 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 facilities and for parking facilities. Assuming the station site to have a rectangular shape with user specified length and width, the fixed cost for station facilities 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 facilities is linear with respect to the park-and-ride demands, their fixed cost is calculated based on a preset unit cost per parking space. The location-based station cost includes the ROW cost and the earthwork cost. The ROW cost and earthwork cost for at-grade stations and parking facilities can be obtained from a GIS.

2.3.2 Operation Cost Saving

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.

$$N_p = \frac{\sum_i \sum_j D_{ijp}}{N_c D_c}, \quad (27)$$

where D_c is the average number of passengers a train car can carry, N_c is the number of cars per train. Assuming the train service is only provided on workdays, the annual energy costs are

$$A_e = (52 \times 5) \sum_p N_p \times E_R \times \frac{K_e}{10^6}, \quad (28)$$

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

$$A_m = (52 \times 5) \sum_p \sum_i \sum_j D_{ijp} L_{i,j} \times \frac{K_O}{10^6}, \quad (29)$$

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

$$A_p = (52 \times 5) \sum_p \sum_i \sum_j D_{ijp}^p L_{i,j}^p \times \frac{K_a}{10^6}, \quad (30)$$

where D_{ijp}^p is the number of park-and-ride trips from Traffic Analysis Zone (TAZ) i to j in period p , $L_{i,j}^p$ is the auto travel distance for D_{ijp}^p (mile), K_a is the unit operation cost for auto (\$/passenger-mile). The original auto operation cost for the rail riders is

$$A_a = (52 \times 5) \sum_p \sum_i \sum_j D_{ijp}^r L_{i,j}^a \times \frac{K_a}{10^6}, \quad (31)$$

where D_{ijp}^r 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

$$A_o = A_a - (A_e + A_m + A_p). \quad (32)$$

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

$$S_o = A_o \frac{(1+r)^{n_a} - 1}{r(1+r)^{n_a}}. \quad (33)$$

2.3.3 User Cost Saving

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

$$A_U = (52 \times 5) \sum_p \sum_i \sum_j D_{ijp}^r \times \left(T_{i,j,p}^a \times \frac{1}{3600} \times \frac{K_U^a}{10^6} - T_{i,j,p}^r \times \frac{1}{3600} \times \frac{K_U^r}{10^6} \right), \quad (34)$$

where $T_{i,j,p}^a$ is the travel time by auto from TAZ i to TAZ j in time period p (s), $T_{i,j,p}^r$ is the travel time by rail from TAZ i to TAZ j in time period p (s), K_U^a is the unit user cost for auto (\$/passenger-hour), K_U^r is the unit user cost for rail (\$/passenger-hour).

The present value of user cost saving is

$$S_U = A_U \frac{(1+r)^{n_a} - 1}{r(1+r)^{n_a}}. \quad (35)$$

3 Case Study

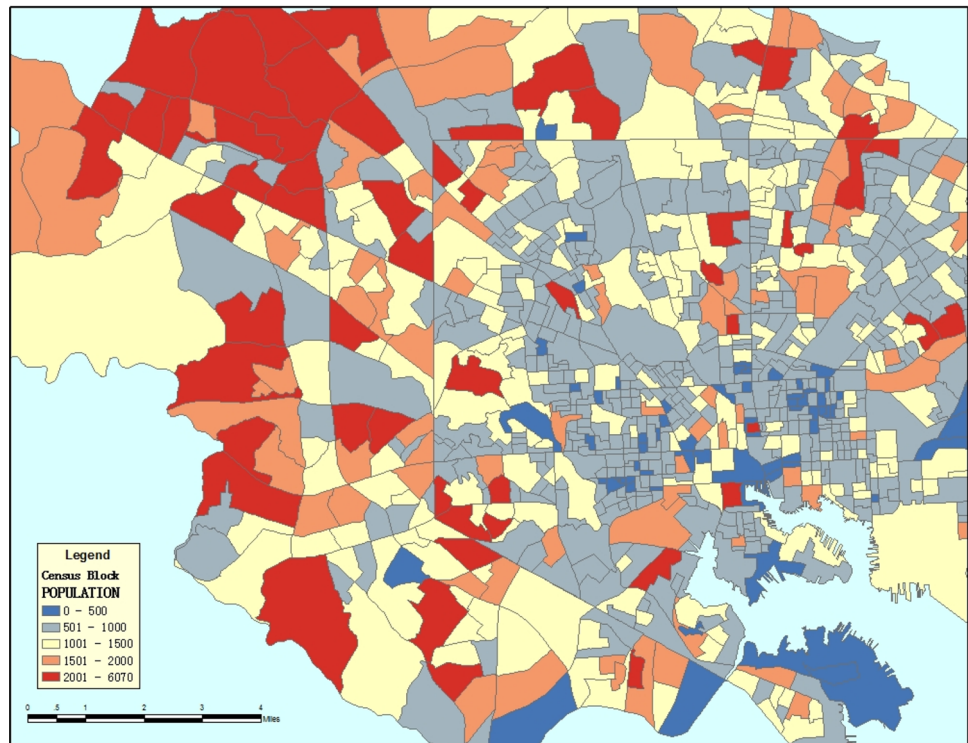
Using Baltimore City as the study area, this case study aims to illustrate the data preparation procedures of the proposed concurrent station location and alignment optimization model, and to demonstrate its effectiveness compared to the sequential optimization methodology where stations are first selected and alignment is then designed between these selected stations.

3.1 Data Preparation for the Proposed Concurrent Optimization Model

3.1.1 Data for Candidate Stations

Following the aforementioned procedures, a grid layer inside the study area is created 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 are 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 overlaid with the land use layer in GIS. All the grids intersecting with those restricted zones are identified as infeasible grids for railway stations.
- Census block data
Census data are obtained from the United States Census Bureau. Year 2000 census data for Baltimore City and Baltimore County are used in the case study, as shown in Fig. 5.

Fig. 5 Census block data

The stations grid cells which attract high population or high employment are considered as potential station locations. As noted earlier, pedestrian-oriented stations and park-and-ride stations have different catchment area dimensions 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 the 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, as shown in Fig. 6.
- **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 cell. The candidate station locations should avoid the high ROW cost grid cells.

After applying the proposed procedures for generating the candidate pool of potential rail transit stations, 52 candidate pedestrian-orientated station locations and 10 candidate park-and-ride station locations are found, as shown in Fig. 7.

3.1.2 Data for Estimating Railway Travel Demand

The proposed concurrent optimization method applies a nested logit mode choice model to calculate mode choice for personal trips and outputs the following four trip tables:

- Mode 1: Drive alone
- Mode 2: High occupancy 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 includes the following variables:

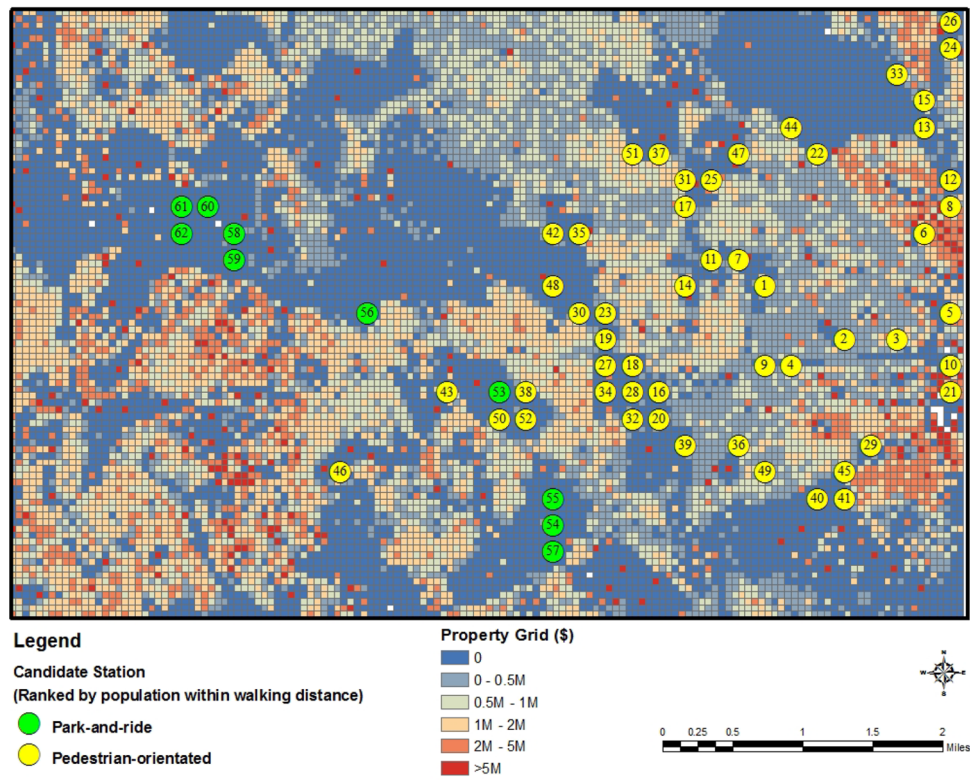
- T_{ijm}^{IN} —Travel time in the vehicle or train from TAZ i to j for mode m
- T_{ijm}^{OUT} —Travel time outside of the vehicle or train from TAZ i to j for mode m
- T^w —Waiting time or headway of the train at the boarding station
- C_{ijm} —Cost of mode m (gas, parking, and ticket) from TAZ i to 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 [27]. TAZ data contain zone-related information, such as population,

Fig. 6 AADT data



Fig. 7 Candidate station locations



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 [27] 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. Then the travel time and distance by train between any two stations are computed via an iterative process based on vehicle dynamics [21].

All of the above data are used to calculate the variables in the utility function. For mode m ,

$$U_{ijm} = (a_m^0 + a_m^1 T_{ij2}^{IN} + a_m^2 T_{ij2}^{OUT} + a_m^3 T_{ij}^w + a_m^4 C_{ijm} + a_m^5 S_{ijm} + a_m^6 A_i + a_m^7 B_j + a_m^8 E_j), \tag{36}$$

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 adapts the values in Table 1 for constants, coefficients, and correlations from BMC model [27]. It is noted that the original model used different sets of parameters depending on the trip type (home-based, work-based, other-based) 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 for examining the effectiveness of the proposed concurrent station and alignment optimization model.

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.

3.2 Model Results

To test the effectiveness of the proposed concurrent station location and alignment optimization model, the following two optimization methods are examined:

- *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.

3.2.1 Comparison of the Two Optimization Methods

Table 2 presents the optimization results of the two-stage optimization and concurrent optimization.

Compared to the two-stage optimization, the concurrent optimization significantly reduces the total cost from 41.8 to 17.6 M. Concurrent optimization reduces the rail line’s passenger trips by 21.8 % compared to the two-stage optimization. By compromising in the passenger attraction, concurrent optimization reduces the travel time from 11.8 to 10.8 min, shortens the track length from 6.7 to 5.8 miles, decreases the initial cost from 74.3 to 54.1 M, and also decreases the operation and user cost for about 4.0 M. The numerical results for this case demonstrate the advantage of concurrent optimization over the two-stage optimization.

Table 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	

Table 2 Comparison of two-stage optimization and concurrent optimization

	Two-stage	Concurrent	Difference (%)
Number of stations	5	5	0
Total daily passengers	4879	3814	−21.8
Travel time (min)	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 and user cost saving (M)	32.5	36.5	12.3
Total cost (M)	41.8	17.6	−57.9

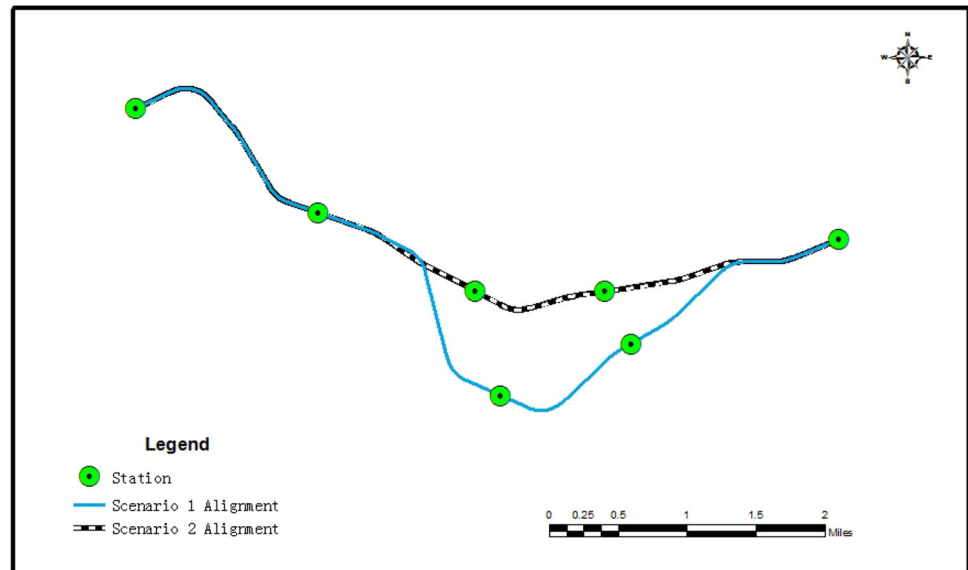
Fig. 8 Optimized station locations and alignments

Figure 8 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 sections of the alignments are shifted to connect different selected stations.

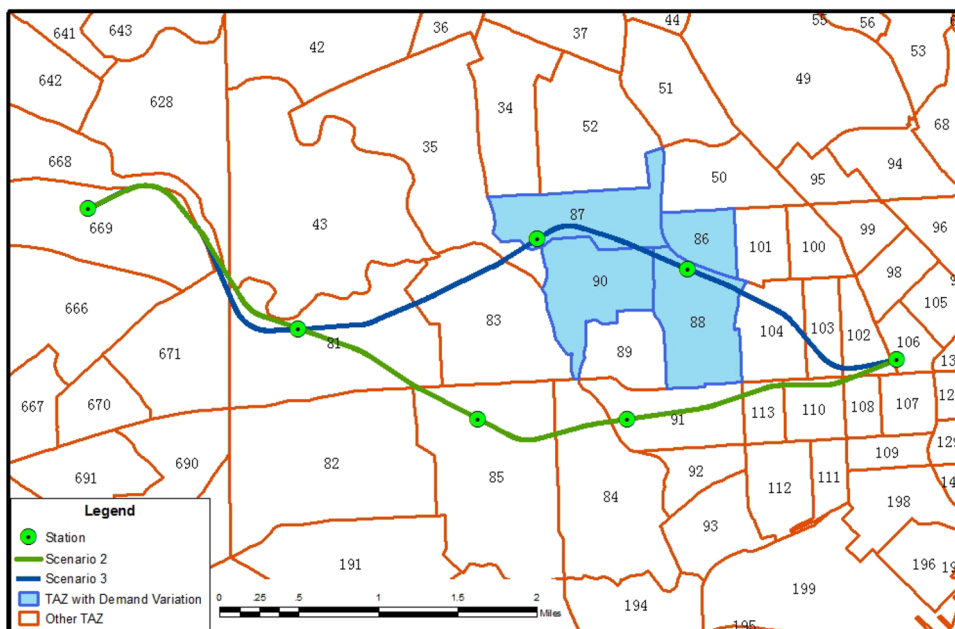
3.2.2 Impact of Demand Variation on Optimization Results

This section examines 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 Fig. 9. Figure 9 also compares the optimized station locations and alignments for the original and the adjusted demand distributions.

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 toward 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 incur an initial cost increase of 7.7 M from 54.1 to 61.8 M. However, the savings in operation and user cost over the system's entire life cycle increase by 91.4 M for the shifted station locations, as they attract more than twice of the original

Fig. 9 Impact of demand variation on optimization results



demand by directly serving the TAZs with higher demand. The results show that the algorithm is effective in adapting the demand patterns and can concurrently optimize the station locations and alignment accordingly.

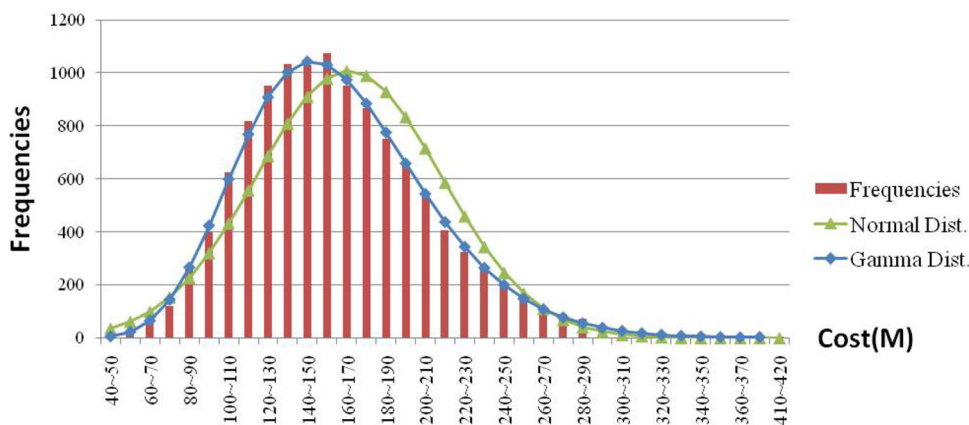
3.2.3 Statistical Test of Solution Goodness

This case study applies a statistical method from Jong and Schonfeld [13] to test the effectiveness of the proposed algorithm in finding near-optimal solutions. This procedure is a sampling process. 50,000 random solutions are generated, of which 23.7 % (11,870) are feasible. The average cost of the feasible solutions is 161.1 M, and standard deviation is 46.9 M. The least cost of any random solution is 41.5 M, which is very far above the best solution (17.6 M) found with our proposed algorithm.

We also use the Gamma and Normal distributions to fit the cost distribution from the feasible random solutions, as shown in Fig. 10, with R^2 values of 0.99 and 0.91, respectively.

The optimized cost from the proposed concurrent optimization model in scenario 2 is 17.6 M, which is better than 99.89 % of solutions in the fitted Normal distribution, and close to 100 % of solutions in the fitted Gamma distribution, but far better than any of the random feasible solutions. This indicates that neither the Normal nor Gamma distributions fit well the extreme low-cost end of the randomly generated solutions. This test shows that although we cannot guarantee a global optimum using the proposed genetic algorithm (or other such metaheuristic algorithms), the optimized solution is extremely good compared to the other solutions in the search space.

Fig. 10 Statistical test of solution goodness



4 Summary

This paper 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 consideration of various site requirements regarding topological features, accessibility to the existing roadway network, and land availability. These candidates are then selected along with the alignment between each pair of neighboring stations using the concurrent optimization model to minimize the total system cost while satisfying station selection and track geometry constraints.

The proposed methodology demonstrates how massive amounts of geographic information can be processed and employed within an optimization framework in planning an urban rail transit system in real-world practice. In particular, it demonstrates how route alignments and station locations can be jointly optimized in continuous space (rather than on an abstracted graph) using whatever relevant geographic information is available. The optimization objectives and constraints used in this paper can be changed without drastically changing the modeling approach presented here. The comprehensive cost evaluation framework addresses the essential trade-off between minimizing capital cost and maximizing savings in operation and user costs by shifting trips from auto to rail. The embedded GA-based heuristic can concurrently optimize the decision variables for station locations, station types, and track alignments connecting each pair of neighboring stations. The case study demonstrates the applicability of the proposed concurrent optimization model and its advantages over the previously used two-stage optimization.

The proposed methodology is expected to help transit planners generate viable alternatives for a single rail transit line effectively and efficiently. Future research could consider the complex interactions between the rail transit line and the existing surface road network, provide the ability to repair infeasible alignments by relaxing some geometric constraints (e.g., by reducing design speeds), improve the computational performance with a distributed genetic algorithm operated on parallel processors, and address the problem of optimizing networks of rail transit lines with feeder bus lines.

Acknowledgments This research was partly funded by the USDOT University Transportation Center Program through the National Center for Strategic Transportation Policies, Investments, and Decisions at the University of Maryland.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use,

distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Shaw JFB, Howard BE (1981) Comparison of two integration methods in transportation routing. *Transp Res Rec* 806:8–13
2. Shaw JFB, Howard BE (1982) Expressway route optimization by OCP. *J Transp Eng* 108(3):227–243
3. Goh CJ, Chew EP, Fwa TF (1988) Discrete and continuous models for computation of optimal vertical highway alignment. *Transp Res Part B* 22(6):399–409
4. Chew EP, Goh CJ, Fwa TF (1989) Simultaneous optimization of horizontal and vertical alignments for highways. *Transp Res Part B* 23(5):315–329
5. Moreb AA (1996) Linear programming model for finding optimal roadway grades that minimize earthwork cost. *Eur J Oper Res* 93(1):148–154
6. ReVelle CS, Whitlatch EE, Wright JR (1997) *Civil and environmental systems engineering*. Prentice Hall, New Jersey
7. Parker NA (1977) Rural highway route corridor selection. *Transp Plan Technol* 3:247–256
8. Trietsch D (1987) A family of methods for preliminary highway alignment. *Transp Sci* 21(1):17–25
9. Trietsch D (1987) Comprehensive design of highway networks. *Transp Sci* 21(1):26–35
10. Fwa TF (1989) Highway vertical alignment analysis by dynamic programming. *Transp Res Rec* 1239:1–9
11. Nicholson AJ, Elms DG, Williman A (1976) A variational approach to optimal route location. *Highw Eng* 23:22–25
12. Jong JC (1998) *Optimizing highway alignments with genetic algorithms*. Ph.D. Dissertation, University of Maryland, College Park
13. Jong JC, Schonfeld P (2003) An evolutionary model for simultaneously optimizing three-dimensional highway alignments. *Transp Res Part B* 37:107–128
14. Jha MK (2000) *A geographic information systems-based model for highway design optimization*. Ph.D. Dissertation, University of Maryland, College Park
15. Jha MK, Schonfeld P (2004) A highway alignment optimization model using geographic information systems. *Transp Res Part A* 38:455–481
16. Jha MK, Schonfeld P, Jong JC, Kim E (2006) *Intelligent road design*. WIT Press, Southampton
17. Kim E (2001) *Modeling intersections & other structures in highway alignment optimization*. Ph.D. Dissertation, University of Maryland, College Park
18. Kim E, Jha MK, Schonfeld P (2004) Intersection construction cost functions for alignment optimization. *J Transp Eng* 130(2):194–203
19. Kang MW, Schonfeld P, Jong JC (2007) Highway alignment optimization through feasible gates. *J Adv Transp* 41(2):115–144
20. Jha MK, Schonfeld P, Samanta S (2007) Optimizing rail transit routes with genetic algorithms and GIS. *J Urban Plan Dev* 133(3):161–171
21. Lai X, Schonfeld P (2012) Optimization of rail transit alignments considering vehicle dynamics. *Transp Res Rec* 2275:77–87
22. Kim M, Schonfeld P, Kim E (2013) Comparison of vertical alignments for rail transit. *J Transp Eng* 139(2):230–238
23. Kim M, Schonfeld P, Kim E (2014) Simulation-based rail transit optimization model. *Transp Res Rec* 2374:143–153
24. Lai X, Schonfeld P (2010) *Optimizing Rail Transit Alignments Connecting Several Major Stations*. Presented at the 89th TRB Annual Meeting, Washington DC
25. Hay WW (1982) *Railroad engineering*. Wiley, New York

-
26. Laporte G, Mesa JA, Ortega FA (2000) Invited review: optimization methods for the planning of rapid transit systems. *Eur J Oper Res* 122:1–10
 27. Baltimore Metropolitan Council (2004) Baltimore Region travel demand model for base year 2000 Task Report 04-01, January 2004