

Optimal Location of Subway Stations: A Case Study on Tehran Subway

Hasti Jafari¹, Masoud Yaghini^{2*}

¹School of Management, University of Tehran, Tehran, Iran ²School of Railway Engineering, Iran University of Science and Technology, Tehran, Iran

ARTICLE INFO	A B S T R A C T			
Article history:	This research is concerned with finding the optimal location of new			
Received: 17.05.2019	stations along the rail transportation network. This is performed by			
Accepted: 23.07.2019	proposing a mathematical model. Two simultaneous effects on railway			
Published: 24.12.2019	demand points (users) are investigated, namely, the time effect and the covered population which is achieved by constructing new stations. The			
Keywords:	improved accessibility of demand points to rail network is considered as the positive one and the negative effect concerns the increased travel time			
Railway station setting	induced by additional stops at newly constructed stations. Two objective			
Station accessibility	functions are considered including the saved travel time and the covered population. The proposed model is also used for a case study on Tehran subway network and the results are presented.			
Population coverage				
Travel time optimization	2 I			

1. Introduction

To design the railway network, the fundamental issue is to find the optimal location of railway stations to improve the passenger satisfaction and the demand of railway systems. When increasing the number of new stations, from the passenger (demand) point of view, two effects need to be investigated simultaneously. The first point to think of is that the passenger tendency to use the rail networks will be increased if the distance to reach the networks is decreased. The second point of concern is the increased travel time at new stations. On the other hand, the new stations should be located in such a way that maximizes the number of people that the rail network can cover. Since late 1960s [1], the rail transport stations' locality has always been an interesting and challenging issue. In 1968 the total cost of locating the stations was minimized by assuming the uniform distribution of passenger travel time [2]. Locating the stations based on a Geographical Information System (GIS) of demand points was considered in [3-5]. A model was developed by Samanta and Jha based on demographic and geographical data and minimized the total cost of station location by using the GA algorithm [4]. The best location of new local stations was determined in four defined phases: site selection, catchment definition, capacity analysis, and service planning. The cost and benefit analysis were estimated on a case study for new local railway stations in UK [5]. A continues stop location problem was considered by a genetic algorithm to minimize the average door-to-door travel time for all customers [6]. Demand coverage and number of stations are factors for locating the stations. Therefore, by considering the prefixed number of stations, the cover of each station was maximized in [7]. Minimizing the number of stops and the demand coverage was considered in a network that has two lines intersecting each other [8].

The railway demand was modeled to develop an indicator for railway accessibility in

[9-11]. The best locations of high-speed rail stations were determined using a Multi Criteria Decision Analysis (MCDA) approach in [12] and the saved travel cost was maximized. In addition, the location problem of high-speed railway stations was considered in the Aveiro city to determine the best locations, by considering the accessibility of different modes of transportation [13]. The location of rail infrastructure was determined by a demand and supply model to locate the lines and stations [14]. Some new analytical approaches were used for locating the station in [15] and [1]. The optimal location of Mashhad railway stations was investigated by a hierarchy and data envelopment analysis model in [15]. A study was conducted to determine the location of stations, which were rated at different operating conditions [1]. The new station location problem on existing rail corridor and new junctions of road network, was modeled in [16]. In 2018, He et al. proposed a model by considering the driving range of electric vehicles for finding the locations of charging stations optimally [17]. In 2019, a model was developed for finding the locations of railway alignment and stations optimally [18]. Mahmoudabadi and Lotfizadeh Jad solved the problem of hob location by proposing a mathematical model. The oriented and destinations nodes considered by using the Gravity model [19]. The quality and rank of each supplier in the supply chain was estimated in Nouhi Tehrani and Bozorgi-Amiri by the usage of experts options. The allocation of material producers to each suppliers was calculated by a mix linear programming [20]. The various methods of developing the underground parking spots in the 7th line of Tehran subway was compared by the analytic hierarchy process [21]. Ahadi and Rashed measured the efficiency of Tehran subway stations by using data envelopment analysis (DEA) [22].

None of the above researches provided a mathematical model to locate the stations optimally based on the covered population and passenger satisfaction. In the current study, a new integer optimization model is presented to determine the optimal location of stations along the rail networks by maximizing the customer's satisfaction and the covered population of the network.

2. Problem Statement

The features of the optimization model presented in this article are similar to Hamacher et al. in 2001 [6]. However, the addition in the present research is the offering of a mathematical model. The objectives of this model are to determine the setting of stations in order to maximize the saved travel time and the network population coverage. To serve the purpose, an optimization model is used.

2.1. Time Model

Time based model consists of two main parts including the distances from demand points to rail network or the time of reaching to the stations. The passenger travel time will be reduced when the new stations are constructed. Consequently, the accessibility to the network will be increased. Travel time for passenger who sits on the train and passes the new stations will be increased due to the additional stops.

The positive effect of time, can be investigated by increasing the user accessibility (demand points) when the stations are constructed. Therefore, the travel time of demand points can be reduced. This reduction of distance is illustrated by L_i - T_i , as L_i indicates the least distance from the points of demand to existing stations and T_i shows the distance among the points of demand to both new and existing stations. The reduced travel time can be calculated by using the average travel speed. The reduction of distance is given as a pair of old and new distances of a demand point to reach the nearest station. This is presented in Equation (1), [6]:

$$\begin{cases} \frac{Li-Ti}{4} & \text{if } L_i \leq 1 \text{ (customer walks)} \\ \frac{Li-Ti}{7} & \text{if } 1 < L_i \leq 5 \text{ (customer rides abike)} \\ \frac{Li-Ti}{20} & \text{if } 1 < L_i \leq 5 \text{ (customer uses bus or car)} \end{cases}.$$
(1)

In Equation (1) the average speed is assumed as 5 km/h [6]. Also, for the purposes of this study the average speed of demand points to reach the stations is considered as 5 km/h.

The negative effect concerns the travel time for passengers onboard who sits on the train and has to travel through the new stations that cause elongation of the total travel time.

3. Mathematical Formulation

A multi-objective model is proposed to maximize the saved travel time and the covered population by the network to increase the customer's satisfaction. In this section, the components of the model are explained and the modeling of the problem is proposed.

The network is shown by G = (K, E) including the already exiting stations and the line segments in the plane shown by K and E, respectively.

The sets

$$J = \{j \in \mathbb{R}^2 : j \in S \text{ for some} l = 1...L\},\$$

$$K = \{k \in \mathbb{R}^2 : k \in S \text{ for some} l = 1...L\}$$

consist of candidate and existing station points along the network respectively and S_l represents the line segment l in the network. Set Irepresents two-dimensional coordinates of demand points and each point is shown by index *i*. For large-scale data, the study area is considered as some different zones to predict the travel of demand points.

The number of population in site *i* is shown by P_i . The area is broken into some zones and the demand points are considered by the center of each zone. The train stop time in each station is assumed to be different and is represented by *delay_j*. In the model, *delay_j* shows different delays in each station and *j* represents the index of that station. The goal is to maximize the saved travel time and these delays are considered by a negative coefficient in the objective function.

The number of passengers who sit on the train and pass the newly staged stations is denoted by W_j (index *j* represents each individual station). The number of such stations is shown by *N*. Two-dimensional coordinates of station location (existing and candidate stations) as well as the location of demand points are the other parameters for the model to calculate the Euclidean distance from the demand points to already existing and new stations, which can be shown by d_{ik} and d_{ij} , respectively. The average travel speed of demand point to reach the stations is shown by *V* in the model as a general parameter and in the calculation process. It is considered as 5 km/h. Binary variable x_{ij} is an

assignment variable: $x_{ij}=1$ if and only if the demand point in site *i* uses the new station in point *j* and 0 for otherwise.

3.1. Population Coverage

Maximizing the population coverage by the network is considered as the second objective function in the model that is presented

as
$$\sum_{i\in I}\sum_{j\in J}v_i x_{ij}$$
.

where the binary variable x_{ij} represents that the demand point in site *i* uses the new station in site *j* or not, and the covered population by the network can be calculated by multiplying the population in each area in the corresponding binary variable.

In addition, the demand point is covered by new stations if:

$$d_{ij}x_{ij} \le r, \tag{2}$$

where d_{ij} refers to Euclidean distance among the point of demands to nearest candidate station and r is some given radius determined by experts in the case study.

The proposed objective functions are presented in Equations (3) & (4) with the proper constraints and restrictions offered in Equations (5-16).

In these equations the binary variable y_j corresponds to the location decision variable: $y_j=1$ if and only if the candidate station at site jis constructed and 0 for otherwise. Binary variables e_k and e_j are used to determine distance between the points of demand to the nearest stations. In this model, L_i shows the minimum distance from the points of demand to existing stations and T_i shows the minimum distance from the points of demand to both existing and newly staged stations. $L_i - T_i$ is the reduction of distances from the demand point to the new stations.

It is possible to estimate the minimum distance in between the stations and points of demands by using this model. The location of demand points and stations (including new and existing) are given as two-dimensional pairs. The procedure for the modeling is summarized in Figure 1.



$$e_{j} \in \{0,1\}, \qquad \forall j, \qquad (15)$$

$$e_k \in \{0,1\}, \qquad \forall K . \tag{16}$$

It should be noted that the distances in between the demand points and stations (existing and new) are considered as data inputs and the minimum distance among the stations (including existing and new stations) and demand points are not considered as input data $(L_i \text{ and } T_i)$ and should be calculated in the model (within the first section of the mathematical model). These minimum distances are not considered as prefixes and will be changed for each *i* and *j* points. Using the



Figure 1. The modeling procedure

notation, the problem can be presented by the following integer optimization:

The first objective function is to maximize the saved travel time and to minimize the additional travel time for the passengers that are already in the train, and have to endure additional stops due to newly constructed stations. In the first term of this objective function, the reduced distance from each demand point to reach the network is divided into the average speed of customers to reach the network (5 km/h) and is multiplied by in zone population. The summation is calculated to obtain the reduced travel time for all network customers. In the second term of this objective function, the time of train stops for all passengers who are passing through the stations is minimized (because of the negative coefficient that is used in the objective function).

The units of Equation (3) is "man-hours". In the first section of this objective function, P_i is the number of population in site *i* and the unit of $(\frac{L_i - T_i}{V})$ is hour. The latter term bears the unit of distance's reduction $(L_i - T_i)$ in kilometer which is divided into the average speed (5 kilometer / hour)). Also, the units of the second section in the first objective function is "man-hours".

In the second objective function, the covered population by the network is maximized and more people can reach the network when the new stations are constructed.

Constraint (5) imposes that the distance between the demand points and the new stations should not exceed the covering radius. The minimum distance from the demand points to existing stations can be found by using constraint (6) and maximizing L_i in the first objective function. It means that each passenger (demand point) uses the nearest station of railway system. Constraint (7) enforced the assignment of points to only one constructed station. In the sixth constraint, the summation of the new station variable e_i and the binary variable of existing stations e_k need to be equal to one. The value of the variable T_i is minimized in the first objective function and equals to minimum amount of $d_{i,j}$ when the binary variable e_i equals unity in the seventh constraint. Constraints (8) to (10) are used to determine the minimum distance from the demand points to all of the existing and new

stations. Constraint (11) guarantees that a demand point is not assigned to a new station that was not constructed. Constraint (12) imposes that the maximum number of stations must be respected.

4. Case Study

A real case study is used to exhibit the efficiency of the proposed model. For this case Tehran subway network is selected. Tehran, the capital of Iran with a population of 14 million is located in the north side of the country.

In this case study, there are 129 subway stations operating within the network. As well as, the optimal number of candidate stations of new designed lines (including 6,7,8 and 9, which are selected by feasibility study by the consultant companies) is calculated as 121 stations which are obtained by using the proposed method of this research.

The existing stations are considered along the lines 1, 2, 3 and 4 within the network. The fifth line of Tehran suburban metro lines was opened in the 1998. It provides an inter-city subway link between Sadeghieh in Tehran to Mehrshahr in Karaj.

The city is divided into 621 zones to facilitate urban activities such as residential, commercial, and industrial activities represented by the central coordinate points, which is presented in Figure 2. By solving the model in CPELEX 11 software, the optimal location of the new stations is determined.



Figure 2. Transportation zones in Tehran

4.1. Optimal Location of New Stations

The maximum covering radius for the modeling is set to a comfortable distance of 1000 m which is presented by r in the model.

The maximum number of new stations is set to 121. The network indicators are summarized in Table 1. The coordinate points of existing and new candidate stations are used to determine the minimum distance from the demand points to existing stations (L_i) and all new and existing stations (T_i) .

Number of existing stations	129
Number of candidate stations	121
Covering radius(meter)	1000
Maximum number of stations to construct	121

The problem to be solved has two objective functions. The multi objective optimization Mathematical Programming and the method to solve are described here:

The mathematical form of multi-objective problem can be stated as:

The number of the first and the second constraints are shown by m_1 and m_2 , respectively. n is the number of objective functions [23]. The implied maximization or minimization for the above equation depends on the problem objectives. In Multi-objective Mathematical Programming (MMP), the optimal exact value which is called non-dominated or Pareto optimality cannot be fully obtained. Therefore, what comes out from the optimization is taken as the best possible solution.

The model for the purposes of this research has two objective functions and is considered as a Multi-objective mathematical programming (MMP), thus the ε -constraint method hat is briefly described hereunder can be applied to solve the problem at hand.

Generally, suppose that there are N objective functions in the model including $f_{\mathbf{x}}(\overline{x}), f_{\mathbf{x}}(\overline{x}), \dots, f_{\mathbf{x}}(\overline{x})$. When ε -constraint method is applied, the first objective function should be considered as the main one and the remaining objective functions as the constrains. To solve the MMP of concern, ɛ-constraint is applied by considering the N-1 objective functions as constrains. The range of each objective function is important and need to be optimally determined. The procedure for solving the MMP by using ε -constraint method is summarized in [24]. The number of objective functions and the vector of decision variables are shown by p and \overline{x} , respectively.

$$\begin{aligned} &Max \quad f_{_{1}}(\bar{x}) \\ &s.t. \\ &f_{_{2}}(\bar{x}) \ge \varepsilon_{_{2}}, f_{_{3}}(\bar{x}) \le \varepsilon_{_{3}}, \dots, f_{_{p}}(\bar{x}) \le \varepsilon_{_{p}} \end{aligned} \tag{17}$$

The optimal value of each objective function can be calculated by solving the MMP problem as a single optimization problem to compute the ranges of each objective function [24]. The optimal value of the objective function *i* is presented by $f_i^*(\overline{x_i})$. The other objective functions include:

$$Min F = \begin{bmatrix} f_1(x), f_2(x), ..., f_n(x) \end{bmatrix}^T$$

s.t.
$$g_i(x) < 0 \qquad i = 1, 2, ..., m_1$$

$$h_i(x) = 0 \qquad i = 1, 2, ..., m_2$$

$$by f_1^*(\overline{x}^{*}_i), f_2^*(\overline{x}^{*}_i), ..., f_p^*(\overline{x}^{*}_i).$$
(18)

Both objective functions that are used in this research are for maximizations which are considered as constraints in the mathematical model that are summarized as:

$$Max \qquad f_{1}(x)$$
st.
$$f_{2}(\overline{x}) \ge \varepsilon_{2}, f_{3}(\overline{x}) \ge \varepsilon_{3}, \dots, f_{p}(\overline{x}) \ge \varepsilon_{p}$$
(19)

The lower bound of the constraints can be calculated when the first objective function is not considered in the model. The estimated optimal value of the second objective function is multiplied by variable coefficient a in Table 2 to obtain the lower bounds for the constraints that is presented in the third column of Table 2. The optimum value of the first objective function and the number of the required stations that are presented in the fourth and the fifth columns of Table 2 can be calculated when the second objective function is considered as a constraint. These values that are summarized in Table 2 are acquired by solving the model in

CPLEX software. The optimal number and locations of candidate stations for construction are 95 and the estimated saved travel time of the first objective function is 2826910 hours, indicating the reduced travel time for network demands.

 Table 2. Samples of optimization parameters for the modeling purposes

f_2^*	а	af_2^*	f_1^*	n	
2772104	0.01	27721.04	9.09E+09	93	
2729797	0.02	54595.94	9.51E+09	95	
2742772	0.03	82283.16	8.86E+09	95	
2892579	0.04	115703.2	9.67E+09	95	
2855846	0.05	142792.3	9.31E+09	98	
:	•	•		•	
2817225	0.23	647961.8	9.43E+09	95	
3047526	0.24	731406.2	1.01E+10	96	
2856659	0.25	714164.8	9.32E+09	94	
2685450	0.26	698217	9.29E+09	94	
2900342	0.27	783092.3	1.02E+10	95	
2724639	0.28	762898.9	9.51E+09	99	
f^*I is the first objective function f^*2 is the second objective function <i>a</i> is the optimization parameter <i>n</i> is the estimated number of stations					

The model needs to run with different coefficients to find the first objective function. The finest size of the first objective function is acquired when coefficient a reaches to a size of 0.27; hence this is called the optimized coefficient for the optimization purposes.

Based on the proposed optimization procedure, the optimal number of candidate stations is 95 and the estimated saved travel time for the first objective function is 2826910 hours, indicating the reduced travel time of the network demands by solving the model in CPLEX software. Additionally, the population covered by the new stations equals to 2900342. The corresponding results are presented in Table 3.

As stated, when a new station is constructed, the access time of demand points to reach the rail network will decrease. On the other hand the travel time for people on the train will increase, too. The estimated size of the first objective function in Table 3 (that is 2826910 hours) indicates the tradeoff among the reduced access time and the increased passenger travel time. The large size of the first objective function is considerable and demonstrates the versatility of the proposed modeling method. Based on this set of forecasts, the population covered by the new stations amounts to 2900342.

Table 3. Model optimal estimations

Optimum size of the first objective function (hours)	2826910	
Optimum size of the second objective function	2900342	
The optimized coefficient (<i>a</i>)	0.27	
The lower bound of the second objective function $(a \times f^*2)$	783092.3	
Optimal number of new stations	95	

5. Conclusions

The issue of finding the proper locations for the railway transportation network stations is in need of considerable financial supports. Also, there are variety of models to logically plan for setting up railway stations. This research proposed a mathematical model to serve the said purpose. The objectives of this model are to maximize the saved railway travel time and to cover a greater range of population

The model components are the travel time, the covered population, and the network accessibility. The model is based on the customer satisfaction to compete with the other transportation modes. It needs to minimize the time losses caused by additional stops at newly constructed stations. It also needs to increase the network accessibility achieved by the reduction of passenger travel time.

Within the proposed model, the objective function, includes two components that is the existing passenger added station delays as well as time savings experienced by population residing in catchment area. Different weighing factors are considered as these two groups of people have different sensitivities to such time changes. A very important piece of information that is needed is the population coverage with the tendency to use the rail when a new station is constructed. Such information should be derived from the local or regional travel demand model to make the proposed model more plausible. This is not considered in this research.

A real case study on Tehran subway is used to check for the versatility of the model.

For the future study, a model based on the above description by considering the operations and construction costs can be generated, while the covering radius can be changed dynamically to suit the needs of the population at each traffic zone.

References

[1] N. Semaan, T. Zayed, A stochastic diagnostic model for subway stations, Tunnelling and Underground Space Technology, Vol. 25, (2010), pp. 32-41.

[2] V.R.Vuchic, G.F. Newell, Rapid transit interstation spacing for minimum travel-time, Transportation Science, Vol. 2, (1968), pp. 303-337.

[3] M.W. Horner, T.W. Grubesic, A GIS-based planning approach to locating urban rail terminals, Transportation, Vol. 28, (2001), pp. 55–77.

[4] S. Samanta, M.K. Jha, Identifying feasible locations for rail transit stations, In Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 2063, (2008), pp. 81-88.

[5] S.P. Blainey, J.M. Preston, A GIS-based appraisal framework for new local railway stations and services, Transport Policy, Vol. 25, (2013), pp. 41-51.

[6] H.W. Hamacher, A. Liebers, A. Schöbel, D. Wagner, F. Wagner, Locating new stops in a railway network, Algorithmic Methods and Models for Optimization of Railways, Vol. 50, No. 1, (2001), pp. 13-23.

[7] G. Laporte, J.A. Mesa, F.A. Ortega, Locating stations on rapid transit lines, Computers and Operation Research, Vol. 29, (2002), pp. 741–759.

[8] M.F. Mammana, S. Mecke, D. Wagner, The station location problem on two intersecting

lines, Electronic Notes in Theoretical Computer Science, Vol. 92, No. 1, (2003).

[9] M. Givoni, P. Rietveld, The access journey to the railway station and its role in passengers' satisfaction with rail travel, Transport Policy, Vol. 14, No. 5, (2007), pp. 357-365.

[10] M. Brons, M. Givoni, P. Rietveld, Access to railway stations and its potential in increasing rail use, Transportation Research Part A: Policy and Practice, Vol. 43, No. 2, (2009), pp. 136-149.

[11] G. Debrezion, E. Pels, P. Rietveld, Modelling the joint access mode and railway station choice, Transportation Research Part E: Logistics and Transportation Review, Vol. 45, No. 1, (2009), pp. 270-283.

[12] R. Mateus, J.A. Ferreira, J. Carreira, Multicriteria decision analysis (MCDA): Central Porto high-speed railway station, European, Journal of Operational Research, Vol. 187, No. 1, (2008), pp. 1-18.

[13] Y. Shen, J.d.A. e Silva, L. Martinez, HSR Station location choice and its local land use impacts on small cities: A case study of Aveiro, Portugal, Procedia - Social and Behavioral Sciences, Vol. 111, (2014), pp. 470-479.

[14] M. Yaghini, , H. Jafari, A mathematical model for finding the optimal locations of railway stations, Advances in Railway Engineering, Vol. 3(1), (2015), pp. 1-8.

[15] N. Mohajeri, G.R. Amin, Railway station site selection using analytical hierarchy process and data envelopment analysis, Computers & Industrial Engineering, Vol. 59, No. 1, (2010), pp. 107-114.

[16] F. Perea, J.A. Mesa, G. Laporte, Adding a new station and a road link to a road-rail network in the presence of modal competition, Transportation Research Part B, Vol. 68, (2014), pp. 1-16.

[17] J. He, H. Yang, T.Q. Tang, H.J. Haung, An optimal charging station location model with the consideration of electric vehicle's driving range, Transportation Research Part C: Emerging Technologies, Vol. 86, (2018), pp. 641-654.

[18] H. Pu, H. Zhang, W. Li, J. Xiong, J. Hu, J. Wang, Concurrent optimization of mountain railway alignment and station locations using a distance transform algorithm, Computers &

[DOI: 10.22068/IJRARE.6.2.123

Industrial Engineering, Vol. 127, (2019), pp. 1297-1314.

[19] A. Mahmoudabadi, Y. Lotfizadeh Jad, Developing passenger hub location problem based on origin-destination trips derived by gravity model (Case study on Iran's rail network), International Journal of Railway Research, Vol. 5, No. 2, (2018), pp. 61-70.

[20] Sh. Nouhi Tehrani, A. Bozorgi-Amiri, Modeling a multi-period transportation system for supplier selection and orders allocation problem (Case study: Parchsaz company), International Journal of Railway Research, Vol. 5, No. 1, (2018), pp. 27-37.

[21] M. Gharouni Nik, R. Esfandiari Mehni, F. Astaraki, E. Hadizadeh Raisi, Optimal selection of underground parking lines arrangement: A case study of Tehran subway line 7, International Journal of Railway Research, Vol. 5, No. 1, (2018), pp. 19-25.

[22] Hamid Reza Ahadi, Farnoosh Rashed, Performance evaluation of Tehran subway stations, International Journal of Railway Research, Vol. 3, No. 2, (2016), pp. 39-46.

[23] J. Aghaei, N.H. Shayanfar, Multi-objective electricity market clearing considering dynamic security by lexicographic optimization and augmented epsilon constraint method, Applied Soft Computing, Vol. 11, (2011), pp. 3846– 3858.

[24] A. Esmaeel Nezhad, M.A.E. Javadi, Rahimi, Applying augmented E-constraint approach and lexicographic optimization to solve multi-objective hydrothermal generation scheduling considering the impacts of pumpedstorage units, Electrical Power and Energy Systems, Vol. 55, (2014), pp. 195–204.