Iranian Journal of Operations Research
Vol. 13, No. 1, 2022, pp. 103-122

Developing a capacitated hub location-routing model for the rapid transit network design under uncertainty

M. Fallah-Tafti[1], M. Honarvar[2,*], R. Tavakkoli-Moghaddam[3], A. Sadegheih[4]

This study aims to develop a capacitated hub location-routing model to design a rapid transit network under uncertainty. The mathematical model is formulated by making decisions about the location of the hub and spoke (non-hub) nodes, the selection of the hub and spoke edges, the allocation of the spoke nodes to the hub nodes, the determination of the hub and spoke lines, the determination of the percentage of satisfied origin-destination demands, and the routing of satisfied demand flows through the lines. Capacity constraints are considered in the hub and spoke nodes and also the hub and spoke edges. Uncertainty is assumed for the demands and transportation costs, represented by a finite set of scenarios. The aim is to maximize the total expected profit, where transfers between the lines are penalized by including their costs in the objective function. The performance of the proposed model is evaluated by computational tests and some managerial insights are also provided through the analysis of the resulting networks under various parameter settings.

***Keywords**: Hub Location, Hub and Spoke Network, Rapid Transit Network, Stochastic Optimization, Transfers.*

Manuscript was received on 03/10/2022, revised on 08/02/2022 and accepted for publication on 09/12/2022.

1. Introduction

Hub location problems (HLPs) are usually considered network location problems in which instead of a direct connection between each pair of given origin-destination (OD) nodes, the flows of goods or services are conveyed from the origin node to the destination through the hub nodes. HLPs are widely used in transportation and logistics, telecommunication, and computer networks. Designing a transportation network in the hub and spoke form makes it possible to reduce the cost of transportation by benefiting from the economies of scale between hubs. Efficient use of scarce transportation resources is another immediate benefit of hubs, as they provide the possibility of connecting a large number of OD pairs by using a small number of links.

Rapid transit network (RTN) design problems usually deal with selecting nodes and links from a potential or underlying network to construct several alignments consisting of stations and connections between them. The establishment of rapid transit systems requires a large investment to install stations and links among them. The success of such investment strongly depends on how well those systems are demanded, which is, in turn, dependent on the network design such as the location of stations. Concerning the features of hub models and the large investment needed to build connections in rapid transit systems, using hub structures for designing them seems to be advantageous.

[*] Corresponding Author.
[1] Department of Industrial Engineering, Yazd University, Yazd, Iran, Email: m.fallah@stu.yazd.ac.ir.
[2] Department of Industrial Engineering, Yazd University, Yazd, Iran, Email: mhonarvar@yazd.ac.ir.
[3] School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran, Email: tavakoli@ut.ac.ir.
[4] Department of Industrial Engineering, Yazd University, Yazd, Iran, Email: sadegheih@yazd.ac.ir.

In the classical hub location and also RTN design problems, it is assumed that all problem data is deterministic and available to the decision-maker in time to make the decision. However, this assumption is restrictive and unrealistic in almost all real-life applications, especially while making long-term decisions such as hub location and RTN design where perfect information is mostly unavailable. The decision maker faces a great deal of uncertainty regarding the problem data stemming from several factors, such as population size shifts and unexpected outbreaks of diseases (e.g., COVID-19). For example, the amount of flow between each pair of nodes in cargo and passenger transportation is not deterministic, and the exact values are not available while making the location decisions. Another uncertainty source can be related to transportation costs. The transportation costs may depend on various random factors such as weather and traffic conditions and fluctuate over time. Consequently, any decision without the consideration of related uncertainty or randomness in the data of the optimization problem can decrease the quality of the obtained solution or even make it infeasible in practice. Therefore, it is important to address the data uncertainty and develop models accordingly.

The present study aims to develop a capacitated hub location-routing model for the RTN design problem in the presence of uncertainty. The uncertainty is considered associated with the demands and transportation costs, which are assumed to be captured by a finite set of scenarios, each with some occurrence probability known in advance. A risk-neutral attitude is suggested for the problem, which means that the current value of future assets will be captured by expected values. The problem is modeled through a two-stage stochastic programming framework. Instead of using pre-assigned configurations, the RTN is planned to be based on a general hub structure with stopovers (stations) in the hub and spoke alignments. In fact, what is sought is a hub location model that uses railway rapid transit systems in both the hub-level sub-network (i.e., the network among the hub nodes) and the spoke-level sub-network (i.e., the network that connects the spoke nodes to each other and the hub nodes). In the hub-level sub-network, more efficient (larger and faster) vehicles are used to benefit from the economies of scale. Due to the employment of rapid transit systems in the hub-level and spoke-level sub-networks, the proposed model relaxes some of the common assumptions and properties in classical hub location models. In this regard, direct connections are allowed between spoke nodes, hub and spoke nodes and edges have considerable setup costs, all the hub and spoke nodes and edges have capacity constraints, the hub-level sub-network is not necessarily a complete network, and paths between origin-destination (OD) pairs do not necessarily contain at least one and at most two hubs. One of the important characteristics of the model is that, unlike generic hub models, it provides the possibility not only for direct links between spoke nodes but also for the transshipment of flows at spoke nodes. Given that rapid vehicles usually rout in lines, from the hub network topology point of view, both hub-level and spoke-level sub-networks are considered to be composed of multiple lines. As such, in addition to designing the network, which involves decisions on the location of the hub and spoke nodes and the selection of hub and spoke edges, the goal of the research is to simultaneously determine hub and spoke rapid transit lines and the way of routing OD demand flows through these lines.

HLPs mainly aim to minimize the total cost of a network to satisfy every demand. However, especially in designing RTNs where setup costs are considerable, it may be more advantageous not to serve every demand. Therefore, this research focuses on profit maximization with no need to serve all the demands. In public transportation systems, in addition to the profit of the system from the operator's point of view, the quality of service should also be enhanced to satisfy the passenger's point of view. The number of transfers within a trip is an important decisive attribute for attracting passengers: transferring is annoying and undesirable for passengers. Therefore, to consider this fact, the costs of transfers between lines are included in the model.

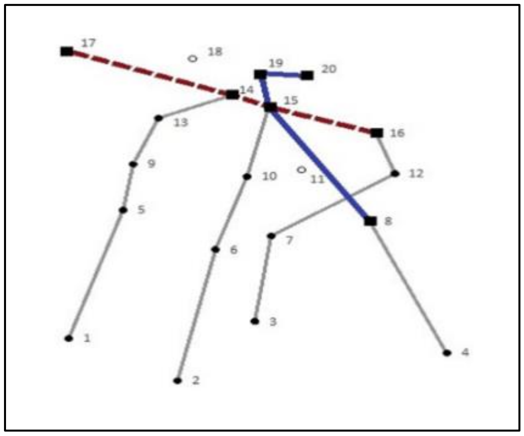


Figure 1. A hub-based RTN

Figure 1 illustrates a hub-based RTN. Hub and spoke nodes are presented with filled squares and circles, and hub and spoke edges are shown with wide and narrow links, respectively. The unfilled circles show the demand centers which are not selected to be serviced. In this study, all the demand centers are considered as the hub and spoke candidates.

The model is evaluated on instances derived from the well-known AP dataset for hub problems, using the CPLEX solver. The computational experiments suggest the effectiveness and efficiency of the proposed stochastic model, and useful insights are provided through the analysis of the resulting networks under various parameter settings. The main contributions of this study are as follows:

a) A mathematical model is introduced to design an RTN based on a hub and spoke model with stopovers (stations) in the hub and spoke alignments.
b) Decisions are made about locating hub and spoke nodes, allocating spoke nodes to hubs, locating hub and spoke edges, constituting hub and spoke transit lines, determining the percentage of OD demand to be served, and determining the way of routing the demand flows through the network.
c) The problem is modeled under uncertainty which is considered associated with the demands and also transportation costs.
d) The capacity constraints are taken into account for all the hub and spoke nodes and the hub and spoke edges.
e) The model incorporates profit maximization with no force assumed to service all the demands and transfer costs are included in the model.

The remainder of the paper is organized as follows. The next section presents a survey of the relevant literature. In Section 3, the problem is defined, and a two-stage mixed-integer linear stochastic programming model is formulated for it. The results of the computational experiments are reported in Section 4. Finally, Section 5 provides some concluding remarks.

2. Literature review

This section presents a review of the relevant literature in the two contexts of hub location and railway rapid transit network design problems.

Five common assumptions underlie most HLPs. They include a) direct connections between spoke nodes are not allowed, b) the costs of edges satisfy the triangle inequality, c) the discount factor α is constant for hub edges, d) hub edges have no setup cost, and e) spoke edges have no setup cost. These assumptions, without any other restrictions, imply two properties. First of all, a hub-level sub-network is a complete network. Secondly, paths between OD pairs do not necessarily contain at least one and at most two hubs. In new approaches, researchers have tried to relax these assumptions to make their models realistic enough to be employed in real-world problems. The relaxation of completeness of hub-level sub-networks has attracted significant attention in recent years; it has not been the case for many real-life networks, for example, in the case of transportation networks where the construction cost of edges is not negligible. In some incomplete hub models, no particular topological structure is considered for the hub-level sub-network, not even being connected [4], while some other models have particular structures including a circle [10], tree [8], star [22] and simple paths or lines [31, 32]. A spoke-level sub-network can also have these structures or other particular structures such as direct connections [1, 29], multi-stops [45], complete sub-graphs [41], and tours [2, 13, 19, 21, 34]. For a review of the particular topological structures of HLPs, the reader is referred to Contreras and O'Kelly [11]. As already mentioned, due to employing rapid transit systems in the hub-level and spoke-level sub-networks, the model in this study relaxes Assumptions 1 and 4-5 and, thus, the corresponding properties. From the perspective of the hub network topology, this study uses a topology of multiple lines for both hub-level and spoke-level sub-networks.

Initially, hub networks were mainly used in air transportation. In recent years, considerable attention has been paid to hub structures for designing multi-modal public transportation networks, especially with rapid transit modes. Some of the studies, in this case, are Chen, et al. [7]; Gelareh and Nickel [15]; Huang, et al. [18]; Kaveh, et al. [20]; Mahéo, et al. [28]; Martins de Sá, et al. [31]; Martins de Sá, et al. [32]; Tavassoli and Tamannaei [40] and Verma, et al. [44]. In these studies, rapid transit modes are usually considered for the hub-level sub-networks, and the spoke-level sub-networks are usually assigned to road vehicles such as cars and trucks [20, 28, 31, 32]. The study of Fallah-Tafti, et al. [14] is the only one that considers RTNs for both hub-level and spoke-level networks. The present study also considers RTN for both hub-level and spoke-level networks and investigate the problem in an uncertain environment considering capacity constraint in the hub and spoke nodes and edges. Studying hub location problems under uncertainty (e.g., fuzzy [35, 43] or stochastic [9, 38, 39] environments) has attracted considerable attention in recent years. In this study, the problem is considered in a stochastic situation and a two-stage stochastic model is developed for it. Studying stochastic hub location problems through a two-stage stochastic modeling framework is not new. The works by Contreras, et al. [9], Correia, et al. [12], and Taherkhani, et al. [38] are worth mentioning in this context. This study aims to develop a stochastic capacitated profit-maximizing hub location model to design an RTN. There are not many studies in the literature focusing on maximization objectives in hub location problems, especially in non-competitive situations [20, 33, 36, 42]. Like other studies (e.g., Oliveira, et al. [33] and Tikani, et al. [42]), this research focuses on profit maximization with no concern about all the demand nodes being necessarily served.

RTN network design problems can be classified into two groups: networks planned from scratch, and additions or extensions of lines in an already functioning network (Canca, et al. [6]). The planned network can consist of one or several lines. To our knowledge, the first mathematical programming model for the general rapid transit network design problem was published by Laporte, et al. [24]. For a more detailed review of the related literature, see Laporte and Mesa [26]. In this context, initial efforts were oriented toward determining a single alignment and locating stations given an alignment (Laporte, et al. [23]). In recent years, however, more realistic cases have been investigated. Some of them (e.g., Gutiérrez-Jarpa, et al. [16] and Gutiérrez-Jarpa, et al. [17]) have dealt with pre-assigned topological configurations to design a network with a single mode of rapid transportation and determined optimal rapid transit lines through pre-assigned corridors. Some others (e.g., Laporte, et al. [25]; López-Ramos, et al. [27], Marín and García-Ródenas [30], and Shushan, et al. [37]) have investigated RTN designs by determining rapid transit lines, without considering pre-assigned corridors. Tactical decisions have also been incorporated into strategic designs, and integrated RTN design and line planning have been addressed in some problems (see e.g., Canca, et al. [5] and Canca, et al. [6]). Instead of using pre-defined configurations, the present study designs an RTN based on general hub and spoke structures with the possibility of stops (stations) for vehicles in the hub and spoke alignments. In addition to the RTN design, rapid lines are determined in the hub-level and spoke-level sub-networks which constitute the first phase of the line planning process.

A summary of the main features of the most related studies is presented in Table 1. For comparison, the characteristics of this study are also presented.

Table 1. Comparison of this paper with the most related studies

Source	Incomplete hub-level network	Spoke-level network		Decisions							Appropriate application in the field of transportation	Transfer costs	Uncertainty	Capacity constraints			Allowing unserved demands
				Location													
		Direct connection of spokes	Transshipment in spokes	Hub nodes	Hub edges	Spoke nodes	Spoke edges	Allocation	Flow routing	Other decisions				Hub nodes	Hub edges	Spoke edges	
Martins de Sá, et al. [31]	✓			✓	✓			✓	✓		public (specially RTN or highway in a hub-level network)						
Verma, et al. [44]	✓	✓	✓	✓				✓	✓	✓	urban bus						
Huang, et al. [18]		✓	✓	✓				✓	✓	✓	bi-modal urban: rail at the hub-level network and bus in the spoke-level network						
Chen, et al. [7]		✓	✓	✓				✓			bi-modal (marine in the hub-level network and rail in the spoke-level network)			✓			
Gutiérrez-Jarpa, et al. [17]						✓	✓		✓		RTN						✓
Canca, et al. [5]						✓	✓		✓	✓	RTN						✓
Canca, et al. [6]						✓	✓		✓	✓	RTN						✓
This paper	✓	✓	✓	✓	✓	✓	✓	✓	✓		**RTN**	✓	✓		✓	✓	✓

3. Model and formulation

In this section, the problem of designing a rapid transit network based on a capacitated hub location-routing model under stochastic uncertainty, which is called a stochastic capacitated hub-based RTN design problem, is modeled. The model is concerned with locating hub and spoke nodes, allocating spoke nodes to hubs, locating hub and spoke edges, constituting hub and spoke transit lines, determining the percentage of OD demands to be served, and determining the way of routing the demand flows through the network. Certain capacity constraints are also postulated in the hub and spoke nodes and edges. Uncertainty is assumed for the demands and transportation costs, represented by a finite set of scenarios, and the objective of the model is to maximize the total expected profit. The main assumptions of the developed mathematical model are as follows:

- Each installed spoke node is located in one and only one spoke line.
- The flows can be routed through the included lines.
- Each spoke node is allocated to at most one hub node.
- To calculate the objective function correctly, all the data related to the costs, revenues and demand flows should be scaled in the same time horizon, for example, yearly or daily.
- Without loss of generality, the underlying or potential graph used as a basis for building the RTN is defined by a set of potential nodes to locate stations and all the possible edges among them.

As a result of the first three assumptions, the flows can change their lines only at the hub nodes. In other words, if the origin and destination of an OD pair are not located in a same line, the corresponding flow has to be routed through the hub-level sub-network. In the following, in order to make the model easier to read, we start by presenting a deterministic version of the problem and afterward we extend it to the stochastic setting.

3.1. A deterministic capacitated hub-based RTN design problem

The modeling procedure of this study is started with presenting a deterministic capacitated hub-based RTN design model, with regard to the defined characteristics of the problem. This deterministic formulation will be later used to develop the stochastic model. The studied deterministic model involves the following sets, parameters and variables.

Sets:

N	Set of the potential nodes to locate the (hub and spoke) stations ($N = \{1, \dots, n\}$).
A	Set of the ordered potential edges among the nodes ($A = N \times N$).
E	Set of edges ($E = \{\{i,j\}: i,j \in N, i < j, (i,j) \text{or } (j,i) \in A\}$).
L^H	Set of possible hub lines (for ease of counting).
L^S	Set of possible spoke lines (for ease of counting).

Parameters:

w_{mn}	Nominal demand flow from node m to node n.
EC_{ij}^h	Construction cost of hub edge $\{i,j\}$.
EC_{ij}^s	Construction cost of spoke edge $\{i,j\}$.
NC_i^h	Construction cost of hub node i.
NC_i^s	Construction cost of spoke node i.
c_{ij}	Nominal transportation cost per unit of flow traveling through edge $\{i,j\}$.
c^t	Transfer cost per transfer and per unit of flow.
d_{ij}	Length of edge $\{i,j\}$.
a	Discount factor for the transportation cost at the hub edges.
r_{mn}	Revenue per unit of flow traveling from node m to node n.

UL^h Upper bound on the number of lines that can cross any hub edge.
HN^{max} Upper bound on the number of the hub nodes per hub line.
SN^{max} Upper bound on the number of the spoke nodes per spoke line.
d^{max} Upper bound on the length of each line.
d^{min} Lower bound on the length of each line.
EC^h Maximum allowed flow in the hub edges.
EC^s maximum allowed flow in the spoke edges (the assumption is $\mathrm{ECap}^s < \mathrm{ECap}^h$).
$NCap^h$ Maximum allowed flow in the hub nodes.
$NCap^s$ Maximum allowed flow in the spoke nodes (the assumption is $NCap^s < NCap^h$).
M A sufficiently large scalar.

Variables:

x_i 1 if node i is selected to be a spoke node; 0, otherwise.
x_{iu} 1 if node i is assigned to hub node u; 0, otherwise ($x_{uu} = 1$ if node u is selected to be a hub node; 0, otherwise).
z_{ij}^h 1 if edge $\{i, j\}$ is selected to be a hub edge; 0, otherwise.
z_{ij}^s 1 if edge $\{i, j\}$ is selected to be a spoke edge; 0, otherwise.
lv_l^h 1 if line l is included as a hub line; 0, otherwise.
lv_l^s 1 if line l is included as a spoke line; 0, otherwise.
x_i^l 1 if node i is selected to be a spoke node of spoke line l; 0, otherwise.
x_{ii}^l 1 if node i is selected to be a hub node of line l; 0, otherwise (If l is a spoke line, then $x_{ii}^l = 1$ means that line l and its nodes are assigned to hub node i).
y_{ijl}^h 1 if hub edge $\{i, j\}$ is selected to be a part of hub line l; 0, otherwise.
y_{ijl}^s 1 if spoke edge $\{i, j\}$ is selected to be a part of spoke line l; 0, otherwise.
v_l 1 if there is at least one hub line other than hub line l included in the network; 0, otherwise.
f_{mnijl}^h Proportion of the flow from m to n pathing through directed hub edge (i, j) in hub line l.
f_{mnijl}^s Proportion of the flow from m to n pathing through directed spoke edge (i, j) in spoke line l.
f_{mn} Proportion of the flow from m to n serviced by the RTN.

The formulation of the deterministic model:

Considering the mentioned notations and assumptions, the formulation of the deterministic hub-based RTN design problem is as follows:

$$\begin{aligned}\max \sum_{m\in N}\sum_{n\in N} r_{mn}.w_{mn}.f_{mn} \\ -\Bigg[\sum_{m\in N}\sum_{n\in N} w_{mn}\sum_{\{i,j\}\in E}\Bigg(\sum_{l\in L^H}\alpha.c_{ij}\big(f_{mnijl}^h + f_{mnjil}^h\big) \\ + \sum_{l\in L^S} c_{ij}\big(f_{mnijl}^s + f_{mnjil}^s\big)\Bigg) - c^t \sum_{m\in N}\sum_{n\in N} w_{mn}\sum_{i\in N: i\neq m,n}\tau_{mni} \\ + \sum_{\{i,j\}\in E}\big(\mathrm{EC}_{ij}^h.z_{ij}^h + \mathrm{EC}_{ij}^s.z_{ij}^s\big) + \sum_{i\in N}\big(\mathrm{NC}_i^h.x_{ii} + \mathrm{NC}_i^s.x_i\big)\Bigg]\end{aligned} \quad (3\text{-}1)$$

s.t.

Network design and line determination constraints:

$$z_{ij}^h + z_{ij}^s \le 1, \quad \{i,j\} \in E \tag{3-2}$$
$$x_i + x_{ii} \le 1, \quad i \in N \tag{3-3}$$
$$x_{iu} \le x_i, \ \ i,u \in N, i \ne u \tag{3-4}$$
$$x_{iu} \le x_{uu}, \ \ i,u \in N \tag{3-5}$$
$$\sum_{u\in N} x_{iu} = x_i + x_{ii}, \ \ i \in N \tag{3-6}$$
$$\sum_{i\in N} x_{ii}^l = \mathrm{lv}_l^s, \ \ l \in L^S \tag{3-7}$$
$$x_i^l \le \sum_{j\in N} x_{jj}^l . x_{ij}, \quad i \in N, l \in L^S \tag{3-8}$$
$$x_{ii}^l \le x_{ii}, \quad i \in N, l \in L^H \cup L^S \tag{3-9}$$
$$x_{ii} \le \sum_{l\in L^H} x_{ii}^l, \quad i \in N \tag{3-10}$$
$$\sum_{l\in L^S} x_i^l = x_i, \quad i \in N \tag{3-11}$$
$$y_{ijl}^h \le z_{ij}^h, \quad \{i,j\} \in E, l \in L^H \tag{3-12}$$
$$y_{ijl}^s \le z_{ij}^s, \quad \{i,j\} \in E, l \in L^S \tag{3-13}$$
$$y_{ijl}^s \le x_i^l + x_{ii}^l, \quad \{i,j\} \in E, l \in L^S \tag{3-14}$$
$$y_{ijl}^s \le x_j^l + x_{jj}^l, \quad \{i,j\} \in E, l \in L^S \tag{3-15}$$
$$y_{ijl}^h \le x_{ii}^l, \quad \{i,j\} \in E, l \in L^H \tag{3-16}$$
$$y_{ijl}^h \le x_{jj}^l, \quad \{i,j\} \in E, l \in L^H \tag{3-17}$$
$$\sum_{l\in LH} y_{ijl}^h \le \mathrm{UL}^h . z_{ij}^h, \quad \{i,j\} \in E \tag{3-18}$$
$$\sum_{i\in N} x_{ii}^l \le \mathrm{HN}^{max} . \mathrm{lv}_l^h, \ \ l \in L^H \tag{3-19}$$
$$\sum_{i\in N} x_i^l \le \mathrm{SN}^{max} . \mathrm{lv}_l^s, \ \ l \in L^S \tag{3-20}$$
$$d^{min} . \mathrm{lv}_l^h \le \sum_{\{i,j\}\in E} d_{ij} y_{ijl}^h \le d^{max} . \mathrm{lv}_l^h, \ \ l \in L^H \tag{3-21}$$
$$d^{min} . \mathrm{lv}_l^s \le \sum_{\{i,j\}\in E} d_{ij}\, y_{ijl}^s \le d^{max} . \mathrm{lv}_l^s, \ \ l \in L^S \tag{3-22}$$
$$\mathrm{lv}_l^h \le \sum_{\{i,j\}\in E} y_{ijl}^h, \quad l \in L^H \tag{3-23}$$
$$\mathrm{lv}_l^s \le \sum_{\{i,j\}\in E} y_{ijl}^s, \quad l \in L^S \tag{3-24}$$

$$\sum_{j:j>i} y_{ijl}^{h} + \sum_{j:i>j} y_{jil}^{h} \leq 2, \quad i \in N, l \in L^{H} \tag{3-25}$$

$$\sum_{j:j>i} y_{ijl}^{s} + \sum_{j:i>j} y_{jil}^{s} \leq 2, \quad i \in N, l \in L^{S} \tag{3-26}$$

$$\mathrm{lv}_{l}^{h} + \sum_{\{i,j\}\in E} y_{ijl}^{h} = \sum_{i\in N} x_{ii}^{l}, \quad l \in L^{H} \tag{3-27}$$

$$\mathrm{lv}_{l}^{s} + \sum_{\{i,j\}\in E} y_{ijl}^{s} = \sum_{i\in N} x_{i}^{l} + \sum_{i\in N} x_{ii}^{l} = \sum_{i\in N} (x_{ii}^{l} + x_{i}^{l}), \quad l \in L^{S} \tag{3-28}$$

$$\sum_{\{i,j\}\in E: i,j\in B} y_{ijl}^{h} \leq |B| - 1, \quad B \subseteq N, |B| \geq 2, l \in L^{H} \tag{3-29}$$

$$\sum_{\{i,j\}\in E: i,j\in B} y_{ijl}^{s} \leq |B| - 1, \quad B \subseteq N, |B| \geq 2, l \in L^{S} \tag{3-30}$$

Connectivity of the network constraints:

$$x_{ii}^{l} \leq \sum_{j:j>i} y_{ijl}^{s} + \sum_{j:i>j} y_{jil}^{s} \leq x_{ii}^{l} + M(1 - x_{ii}^{l}), \quad i \in N, l \in L^{S} \tag{3-31}$$

$$\sum_{l'\in L^{H}: l'\neq l} \sum_{i\in N} x_{ii}^{l}.x_{ii}^{l'} \geq \mathrm{lv}_{l}^{h}.\nu_{l}, \quad l \in L^{H} \tag{3-32}$$

$$\nu_{l} \leq \sum_{l'\in L^{H}: l'\neq l} \mathrm{lv}_{l'}^{h} \leq M.\nu_{l}, \quad l \in L^{H} \tag{3-33}$$

Flow routing constraints:

$$f_{mnijl}^{h} + f_{mnjil}^{h} \leq y_{ijl}^{h}, \quad m,n \in N, \{i,j\} \in E, l \in L^{H} \tag{3-34}$$

$$f_{mnijl}^{s} + f_{mnjil}^{s} \leq y_{ijl}^{s}, \quad m,n \in N, \{i,j\} \in E, l \in L^{S} \tag{3-35}$$

$$\sum_{m\in N} \sum_{n\in N} \sum_{l\in L^{H}} w_{mn} \left(f_{mnijl}^{h} + f_{mnjil}^{h}\right) \leq \mathrm{ECap}^{h}, \quad \{i,j\} \in E \tag{3-36}$$

$$\sum_{m\in N} \sum_{n\in N} \sum_{l\in L^{S}} w_{mn} \left(f_{mnijl}^{s} + f_{mnjil}^{s}\right) \leq \mathrm{ECap}^{s}, \quad \{i,j\} \in E \tag{3-37}$$

$$\sum_{m\in N} \sum_{n\in N} \sum_{J\in N} w_{mn} \left(\sum_{l\in L^{H}} f_{mnjil}^{h} + \sum_{l\in L^{S}} f_{mnjil}^{s} \right) \leq \mathrm{NCap}^{h}.x_{ii} + \mathrm{NCap}^{s}.x_{i}, \quad i \in N \tag{3-38}$$

$$\sum_{j\in N} \left(\sum_{l\in L^{H}} \left(f_{mnijl}^{h} - f_{mnjil}^{h}\right) + \sum_{l\in L^{S}} \left(f_{mnijl}^{s} - f_{mnjil}^{s}\right) \right) \begin{cases} = +f_{mn}, & m,n,i \in N: i = m, m \neq n \\ = -f_{mn}, & m,n,i \in N: i = n, m \neq n \\ = 0, & m,n,i \in N: i \neq m, i \neq n \end{cases} \tag{3-39}$$

Bounding and sign constraints:

$$x_{i}, x_{iu}, z_{ij}^{h}, z_{ij}^{s}, x_{ii}^{l} \in \{0,1\}, \quad i,u \in N, \{i,j\} \in E, l \in L^{H} \cup L^{S} \tag{3-40}$$

$$\mathrm{lv}_l^h, y_{ijl}^h, v_l \in \{0,1\}, \quad l \in L^H, \{i,j\} \in E \tag{3-41}$$
$$\mathrm{lv}_l^s, x_i^l, y_{ijl}^s \in \{0,1\}, \quad l \in L^S, i \in N, \{i,j\} \in E \tag{3-42}$$
$$f_{mn} \in (0,1), \quad m,n \in N \tag{3-43}$$
$$f_{mnijl}^h \in (0,1), \quad l \in L^H, m,n,i,j \in N \tag{3-44}$$
$$f_{mnijl}^s \in (0,1), \quad l \in L^S, m,n,i,j \in N \tag{3-45}$$

Objective function (3-1) maximizes the net profit which is calculated by the subtraction of the total costs, including the transportation costs and the installation costs of the hub and spoke nodes and edges, from the total revenue obtained through satisfying the demands. The revenue is considered as the sum of the ticket price and the government subsidy. Constraints (3-2) and (3-3) guarantee that each installed node and edge can only be a hub or a spoke. Constraints (3-4) and (3-5) are added to the model to define the variable x_{iu} correctly. Constraints (3-6) hold that each installed node is assigned to one and only one hub node. Each hub node is assigned to itself. Constraints (3-7) enforce each included spoke line to be assigned to one and only one hub node. Constraints (3-8) guarantee that spoke lines assigned to a particular hub node can only consist of the spoke nodes which are assigned to that hub node. Constraints (3-9) mean that a hub station is selected to be the station of a line only if it is already built in the network. Constraints (3-9) ensure that a built hub station is located in at least one hub line. Constraints (3-11) mean that a spoke station is selected to be the station of a line only if it is already built in the network and that it is located in one and only one spoke line (i.e., flows can change their line only at the hub nodes). Constraints (3-12) and (3-13) enforce an edge to be included in a line if it is already built in the network. Constraints (3-14) to (3-17) ensure that an edge is built in the network only if the adjacent nodes are already built. Constraints (3-18) impose an upper bound on the number of the lines that can circulate at any hub edge of the network, i.e., they prevent a concentration of lines at hub edges (constraints (3-11) enforce this bound to be equal to one for the spoke edges). Constraints (3-19) and (3-20) postulate a maximum number of nodes for each line and impose that no node or edge can be part of a non-included line. Constraints (3-21) and (3-22) assign maximum and minimum lengths to each line. According to the constraints (3-23) and (3-24), at least one edge must be present in each line. Constraints (3-25) to (3-30) determine the lines' topology. Constraints (3-25) and (3-26) enforce each node of a line to have at most two incident edges. Constraints (3-27) and (3-28) make the number of the edges of a line equal to the number of its nodes minus one. Constraints (3-29) and (3-30) are sub-tour elimination constraints that guarantee that no line contains cycles. Constraints (3-31) guarantee that the hub node corresponding to a specific spoke line is a node of that line and the line is connected to the hub-level sub-network through that hub node. These constraints ensure the connectivity of the spoke-level sub-network to the hub-level sub-network and make sure that all the spoke lines are connected to the hub-level sub-network through their corresponding hub nodes. According to constraints (3-25) and (3-26), one can consider $M = 2$ for the constraints already enumerated. Constraints (3-32) postulate connectivity for the hub-level sub-network and mean that each hub line must share at least one node with at least one other included hub line. Constraints (3-33) are added to the model to define the v_l variables correctly. Constraints (3-34) and (3-35) mean that, if the flow corresponding to an OD pair uses an edge of a line, the edge must already be selected to be a part of that line. Constraints (3-36) and (3-37) address the capacity constraints of the hub and spoke edges. Constraints (3-38) address the capacity constraints of the hub and spoke nodes. Constraints (3-39) relate to flow conservation and ensure that the flow from m to n leaves node m, arrives at node n, and is accounted for whenever a middle node i is used. Finally, constraints (3-40) to (3-45) are bounding and sign constraints.

3.1. The stochastic capacitated hub-based RTN design problem

A stochastic version of the above problem is considered in which uncertainty relates to the demands, w_{ij}s, and the transportation costs c_{ij}s. It is assumed that uncertainty of the problem can be captured by a finite set of scenarios for demands and transportation costs that are denoted by S_w and S_c, respectively. Each scenario, $s_w \in S_w$ and $s_c \in S_c$, is assumed to occur independently with some known probability p_{s_w} and p_{s_c}, respectively. A two-stage modeling framework is considered to model the stochastic problem. The first stage decisions concern the location of the hub and spoke nodes, the selection of the hub and spoke edges, the allocation of the spokes to the hubs, and the determination of the hub and spoke lines. The decisions about the routing of flows through the lines are included in the second stage. Since the random vectors underlying the stochastic problem have a finite support, the second-stage decision variables can be indexed in the scenario sets. In order to do that and with regard to the independence of the random variables, the data concerning flows, which they now are random variables as well, is redefined as follows:

w_{mns_w}	The demand flow from node m to node n under scenario s_w.
c_{ijs_c}	The demand flow from node m to node n under scenario s_c.
$f^h_{mnijls_ws_c}$	The proportion of the flow from m to n pathing through directed hub edge (i,j) in hub line l under scenarios s_w and s_c.
$f^s_{mnijls_ws_c}$	The proportion of the flow from m to n pathing through directed spoke edge (i,j) in spoke line l under scenario s_w and s_c.
$f_{mns_ws_c}$	The proportion of the flow from m to n serviced by the RTN under scenarios s_w and s_c.
$\tau_{mnis_ws_c}$	The proportion of the flow from m to n serviced by the RTN under scenarios s_w and s_c.

The extensive form of the deterministic equivalent model is now written as follows.

$$\begin{aligned}\max \sum_{s_w\in S_w}\sum_{s_c\in S_c}\sum_{m\in N}\sum_{n\in N} & p_{s_w}.p_{s_c}.r_{mn}.w_{mns_w}.f_{mns_ws_c}\\
&-\sum_{s_w\in S_w}\sum_{s_c\in S_c}\sum_{m\in N}\sum_{n\in N} p_{s_w}.p_{s_c}.w_{mns_w}\sum_{\{i,j\}\in E}\Bigg(\sum_{l\in L^H}\alpha.c_{ijs_c}\big(f^h_{mnijls_ws_c}\\
&+f^h_{mnjils_ws_c}\big)+\sum_{l\in L^S}c_{ijs_c}\big(f^s_{mnijls_ws_c}+f^s_{mnjils_ws_c}\big)\Bigg)\\
&-c^t\sum_{s_w\in S_w}\sum_{s_c\in S_c}\sum_{m\in N}\sum_{n\in N} p_{s_w}.p_{s_c}.w_{mns_w}\sum_{i\in N:i\neq m,n}\tau_{mnis_ws_c}\\
&-\sum_{\{i,j\}\in E}\big(FCh_{ij}.zh_{ij}+FCs_{ij}.zs_{ij}\big)-\sum_{i\in N}(ch_i.x_{ii}+cs_i.x_i)\end{aligned} \tag{3-46}$$

s.t.

(3-2)-(3-33), (3-40)-(3-42)

$$f^h_{mnijls_ws_c}+f^h_{mnjils_ws_c}\le y^h_{ijl},\quad m,n\in N,\{i,j\}\in E,l\in L^H,s_w\in S_w,s_c\in S_c \tag{3-47}$$

$$f^s_{mnijls_ws_c}+f^s_{mnjils_ws_c}\le y^s_{ijl},\quad m,n\in N,\{i,j\}\in E,l\in L^S,s_w\in S_w,s_c\in S_c \tag{3-48}$$

$$\sum_{m\in N}\sum_{n\in N}\sum_{l\in L^H} w_{mns_w}.\big(f^h_{mnijls_ws_c}+f^h_{mnjils_ws_c}\big)\le \mathrm{ECap}^h,\quad \{i,j\}\in E,s_w\in S_w,s_c\in S_c \tag{3-49}$$

$$\sum_{m\in N}\sum_{n\in N}\sum_{l\in L^S} w_{mns_w}.\big(f^s_{mnijls_ws_c}+f^s_{mnjils_ws_c}\big)\le \mathrm{ECap}^s,\quad \{i,j\}\in E,s_w\in S_w,s_c\in S_c \tag{3-50}$$

$$\sum_{m\in N}\sum_{n\in N}\sum_{j\in N} w_{mns_w}\left(\sum_{l\in L^H} f^h_{mnjils_ws_c} + \sum_{l\in L^S} f^s_{mnjils_ws_c}\right) \leq \mathrm{NCap}^h . x_{ii} + \mathrm{NCap}^s . x_i, \quad i \in N, s_w \in S_w, s_c \in S_c \tag{3-51}$$

$$\sum_{j\in N}\left(\sum_{l\in L^H}\left(f^h_{mnijls_ws_c} - f^h_{mnjils_ws_c}\right) + \sum_{l\in LS}\left(f^s_{mnijls_ws_c} - f^s_{mnjils_ws_c}\right)\right) \begin{cases} = +f_{mns}, & m,n,i \in N: i = m, m \neq n, s_w \in S_w, s_c \in S_c \\ = -f_{mns}, & m,n,i \in N: i = n, m \neq n, s_w \in S_w, s_c \in S_c \\ = 0, & m,n,i \in N: i \neq m, i \neq n, s_w \in S_w, s_c \in S_c \end{cases} \tag{3-52}$$

$$f_{mns_ws_c} \in (0,1), \quad m,n \in N, s_w \in S_w, s_c \in S_c \tag{3-53}$$

$$f^h_{mnijls_ws_c} \in (0,1), \quad l \in L^H, m,n,i,j \in N, s_w \in S_w, s_c \in S_c \tag{3-54}$$

$$f^s_{mnijls_ws_c} \in (0,1), \quad l \in L^S, m,n,i,j \in N, s_w \in S_w, s_c \in S_c \tag{3-55}$$

As already mentioned, the first-stage problem consists of defining the network infrastructure by locating the hub and spoke nodes, selecting the hub and spoke edges, allocating the spoke nodes to the hub ones, as well as determining the hub and spoke lines. The objective function (3-46) represents the corresponding profits and also includes the expected profits for routing the flows through the lines. We note that in the second-stage problem, variables y^h_{ijl} and y^s_{ijl} are fixed. Therefore, we are facing a routing problem with predetermined (hub and spoke) lines.

3.2. Linearization of the models

To linearize constraints (3-8), the binary variables

$$z^l_{ij} = x^l_{jj} . x_{ij}, \quad i,j \in N, l \in L^S,$$

are added to the model, and these constraints are replaced with the following ones.

$$x^l_i \leq \sum_{j\in N} z^l_{ij}, \quad i \in N, l \in L^S \tag{3-56}$$

$$z^l_{ij} \leq x^l_{jj}, \quad i,j \in N, l \in L^S \tag{3-57}$$

$$z^l_{ij} \leq x_{ij}, \quad i,j \in N, l \in L^S \tag{3-58}$$

$$x^l_{jj} + x_{ij} - 1 \leq z^l_{ij}, \quad i,j \in N, l \in L^S \tag{3-59}$$

To linearize constraints (3-32), the binary variables

$$z^{ll'}_i = x^l_{ii} . x^{l'}_{ii}, \quad i \in N, l, l' \in L^H : l' > l$$
$$z^l = \mathrm{lv}^h_l . \nu_l, \quad l \in L^H$$

are added to the model, and these constraints are replaced with the following ones.

$$\sum_{l'\in L^H: l' > l}\sum_{i\in N} z^{ll'}_i + \sum_{l'\in L^H: l > l'}\sum_{i\in N} z^{l'l}_i \geq z^l, \quad l \in L^H \tag{3-60}$$

$$z^{ll'}_i \leq x^l_{ii}, \quad i \in N, l, l' \in L^H : l' > l \tag{3-61}$$

$$z_i^{ll'} \leq x_{ii}^{l'} \quad i \in N, l, l' \in L^H : l' > l \tag{3-62}$$

$$x_{ii}^{l} + x_{ii}^{l'} - 1 \leq z_i^{ll'}, \quad i \in N, l, l' \in L^H : l' > l \tag{3-63}$$

$$z^l \leq \mathrm{lv}_l^h, \quad l \in L^H \tag{3-64}$$

$$z^l \leq v_l, \quad l, l' \in L^H : l' > l \tag{3-65}$$

$$\mathrm{lv}_l^h + v_l - 1 \leq z^l, \quad l \in L^H \tag{3-66}$$

4. Computational Experiments

This section presents the results of the computational experiments performed to assess the performance of the proposed stochastic capacitated hub-based RTN design model and to analyze the resulting hub RTN networks. The main purposes are (i) to check whether the proposed model can help to find an optimal or near-optimal solution in a reasonable CPU time; (ii) to evaluate the relevance of considering a stochastic modeling framework for the problem investigated; and (iii) to measure the sensitivity of the model to variations in some important parameters.

The stochastic capacitated hub-based RTN design model (defined by equations (3-46)-(3-55), (3-2)- (3-7), (3-9)-(3-31), (3-33), (3-40)-(3-42), (3-56)-(3-66)) was implemented using IBM ILOG Concert Technology with IBM CPLEX 12.10. The experiments were conducted on a personal laptop with the Intel® Core™ i5-8250U CPU @ 1.60-1.80 GHz processor and Windows 10 with 8 GB of RAM memory.

3.2. Test data

The AP dataset was used to perform the computational experiments. In order to calculate the data related to costs, revenues and demand flows on the same scale, the objective function was calculated in a time horizon of H years. To this end, the nominal demand flow, i.e., the nominal number of passengers, from i to j in H years, which is denoted by w_{ij}, was calculated as $w_{ij} = w_{ij}^{AP} / \sum_{i,j \in N} w_{ij}^{AP} \times t^{AP} \times H$, where w_{ij}^{AP} is the demand flow of i to j taken from the AP dataset and t^{AP} is the estimated total number of the yearly trips. The distance matrix of the model was considered equal to the distance matrix provided by the dataset. The remaining parameters were set as reported in Table 2. The values in this table were used in all the experiments unless it is explicitly mentioned. In the following, we present the results of the mentioned computational experiments and give some managerial insights.

3.3. Computational experiments

This section presents the results of the computational experiments performed to assess the performance of the proposed model and to analyze the resulting hub RTN networks under different parameter settings and give some managerial insight. For scenario generation, a perturbation level of $[0.5,1.5]w_{ij}$ and $[0.5,1.5]c_{ij}$ is considered for the demand flow and transportation cost between each pair of nodes, respectively, in which $[0.5,1.5]$ represents the interval for the uniform distribution with minimum value 0.5 and maximum value 1.5, and w_{ij} and c_{ij} are the nominal demand flow from node i to node j, obtained from the dataset, and the nominal transportation cost from node i to node j, given in Table 2, respectively. The probabilities of generated scenarios are assumed to be equal.

Table 2. Values of the input parameters for the stochastic capacitated hub-based RTN model

Name	Value	Name	Value
n	10	UL^h	1
$\lvert L^H \rvert$	1	HN^{max}	3
$\lvert L^S \rvert$	2	SN^{max}	5
$\lvert S_w \rvert$	2	d^{max}	$\max d_{ij}$ (km)
$\lvert S_c \rvert$	2	d^{min}	$\min d_{ij}$ (km)
EC_{ij}^h	$7 \times 10^7 d_{ij}(1 + 0.5\upsilon_{\mathrm{ECap}})$ (cur)	ECap^h	$1.75\mathrm{ECap}^s$
EC_{ij}^s	$5 \times 10^7 d_{ij}(1 + 0.25\upsilon_{\mathrm{ECap}})$ (cur)	ECap^s	$\upsilon_{\mathrm{ECap}}.\max \widehat{w}_{ij}, \upsilon_{\mathrm{ECap}} = 1$
NC_i^h	$1.25 \times 10^8(1 + 0.4\upsilon_{\mathrm{NCap}})$(cur)	NCap^h	$2.5\mathrm{NCap}^s$
NC_i^s	$8 \times 10^7(1 + 0.15\upsilon_{\mathrm{NCap}})$ (cur)	NCap^s	$\upsilon_{\mathrm{NCap}}.\max \widehat{w}_{ij}, \upsilon_{\mathrm{NCap}} = 2$
c_{ij}	$100d_{ij}$ (cur)	M	2
c^t	$0.1\bar{c}, \bar{c} = \sum_{i,j} c_{ij}\, n^2$ (cur)	H	30
a	0.7	t^{AP}	10^7
r_{mn}	$\upsilon_r d_{mn}, \upsilon_r = 150$		

To gain an insight into the performance of the proposed model, in the first part of the experiments, its capability to reach an optimal or near-optimal solution in a reasonable CPU time and the relevance of considering the stochastic modeling framework for the problem were investigated. Stochastic problems are usually more difficult to solve than their deterministic counterpart. Therefore, it is important to evaluate the relevance of using the stochastic model instead of working with deterministic versions of the problem. The expected value of the perfect information ($EVPI$) and the value of the stochastic solution (VSS) are often considered to indicate this relevance (Birge and FV [3]). The $EVPI$ represents the difference between the optimal value of the stochastic problem, SP, and the objective value of the so-called wait-and-see solution, WS. Let DE_s denote the optimal value of the deterministic problem associated with scenario $s \in S$. We have $EVPI(\%) = (WS - SP)/SP \times 100$, where $WS = \sum_{s \in S} p_s DE_s$. The percentage of VSS is defined as $VSS(\%) = (SP - EEV)/SP \times 100$, where EEV is the value of the stochastic problem when the first stage variables are fixed according to their optimal values in the deterministic problem resulting from considering the scenario obtained when each random variable is replaced by its expected value. The first part of the experiments were conducted by varying the number of scenarios of the demands and transportation costs. Table 3 to Table 5 present the results obtained for the cases of $|S_w| \in \{1,2,3\}, |S_c| = 1$ (stochastic demands), $|S_w| = 1, |S_c| \in \{1,2,3\}$ (stochastic transportation costs), and $|S_w| \in \{1,2,3\}, |S_c| \in \{1,2,3\}$ (stochastic demands and transportations costs), respectively. For each number of scenarios, the next four columns indicate the corresponding values for the CPU time (s) required to obtain the optimal solution, the optimal expected total profit, the $EVPI(\%)$, and the $VSS(\%)$, respectively. Each reported result in the tables is the average values of three instances.

Table 3. Performance of the model with stochastic demands

Row	$\lvert S_w \rvert$	$\lvert S_c \rvert$	CPU Time	P	$EVPI(\%)$	$VSS(\%)$
1	1	1	144	7.85335e+13	-	-
2	3	1	807	9.35546e+13	0.57	0.00
3	5	1	5683	6.71349e+13	0.60	0.46

Table 4. Performance of the model with stochastic transportation costs

Row	$\lvert S_w \rvert$	$\lvert S_c \rvert$	CPU Time	P	$EVPI(\%)$	$VSS(\%)$
1	1	1	144	7.85335e+13	-	-
2	1	3	1282	6.68219e+13	12.01	0.00
3	1	5	6400	6.46517e+13	13.33	3.34

Table 5. Performance of the model with stochastic demands and transportation costs

Row	$\lvert S_w \rvert$	$\lvert S_c \rvert$	CPU Time	P	$EVPI(\%)$	$VSS(\%)$
1	1	1	144	7.85335e+13	-	-
2	2	2	2709	7.47316e+13	8.88	0.37
3	3	3	24057	7.17788e+13	13.09	0.01

From the tables, one can observe that, not surprisingly, the instances become harder to solve as the number of scenarios increases due to the increase in the number of variables and constraints in the model. It is also observed that the values of the $EVPI(\%)$ and $VSS(\%)$ factors are considered on average (with average values of 8.08% and 0.70%, respectively) for all cases and increase by an increase in the number of scenarios. It can be concluded that knowledge about the future is considered relevant and it is worth solving the stochastic problem, especially when the number of scenarios increases. It also shows that making decisions under uncertainty can change the network structure. Note that the first row of the table presents the results for the deterministic model. Therefore, $EVPI(\%)$ and $VSS(\%)$ values are not reported for it.

To analyze the resulting hub RTN networks under different parameter settings and get insights into the characteristics of the solutions obtained by the model, in the second part of the experiments, the effect of the discount factor, the revenues, and the edges and nodes capacity upper bounds, on the obtained solutions, were analyzed. To this end, the solutions were investigated in terms of the expected total net profit, the expected percentage of the served demands, and the topology of the network to give some managerial insight.

For this part, the experiment related to the second row of Table 5 with $|S_c| = 3, |\, S_c| = 2$ is referred to again. For each case, some instances were solved with the varied values of the parameter under investigation but the fixed values of the other parameters. Table 6 to Table 8 give the results of the experiments conducted to investigate the effect of discount factor, incomes and capacity bounds, respectively. For the investigation of r_{mn}s, the parameter υ_r and the capacity constraints of the parameters υ_{ECap} and υ_{NCap}, given in Table 2, were varied as given in Table 7 and Table 8, respectively. Note that in the case of investigation of capacity bounds the related installation costs of nodes and edges are set according to the values of the υ_{ECap} and υ_{NCap} parameters, as given in Table 2. For each size of the problem, the next four columns of the tables indicate the corresponding values for the parameter under investigation, the total percentage of the satisfied demand, and the total net profit, respectively. Figure 2 presents the resulting network configurations for one of the three instances. The thick line represents the hub, and the narrow ones represent the spoke lines. The nodes that are not connected to the network are the ones that are not serviced.

The results of Table 6 show that the decision-maker can obtain significantly more profit when the discount is higher due to the economies of scale. It is also observed that by increasing the discount, the percentage of the captured demands at first increases and then decreases. It can shows this fact that increasing the discount to some extent can increase the profit due to increasing the captured demands and more than this can only increase the profit due to the discount of the costs and not the increased demand capture. As understood from Table 7 and as expected, an increase in the revenue parameter υ_r leads to a rise in the percentage of satisfied demand and the total net profit. According to Figure 2 ((d), (e), and (f)), it seems that lower revenues give rise to smaller networks with fewer numbers of stations and shorter edges. This is because as revenues and, therefore, satisfied demands decrease, high investment for building a large number of stations and larger edges cannot be profitable. According to Table 8, increasing the capacity bounds also leads to the rise of the percentage of the satisfied demand and the total net profit which can mean that the increased incomes coming from the increased captured demands, due to the increase of the capacity bounds, overcome the related increased nodes and edges installation costs. For this model, increasing discounts, incomes, and capacity bounds all increase the average CPU time. This indicates that when fewer discounts, incomes, or capacity bounds are applied, the instances tend to be easier to solve.

Using the computational results, the effectiveness and efficiency of the proposed stochastic model were validated, and useful insights were provided into the interactions among the different aspects of the studied complex decision problem. From these results, one can conclude that different planning decisions are jointly involved in obtaining good solutions. This shows the importance of considering different aspects and phases of a planning process to obtain good solutions for real-life situations.

Table 6. Solution characteristics for different values of the discount factor.

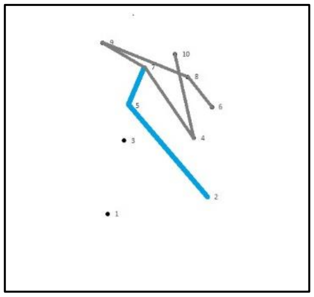
Row	$1-a$	Satisfied demand (%)	Profit value	CPU Time
1	0.1	18.38	6.53685e+13	2709
2	0.3	18.47	7.47316e+13	2709
3	0.5	16.00	8.56125e+13	2770

Table 7. Solution characteristics for different values of the income factor.

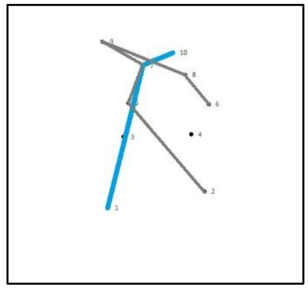
Row	γ	Satisfied demand (%)	Profit value	CPU Time
1	110	15.02	3.62493e+13	358
2	150	18.47	7.47316e+13	2709
3	200	18.51	1.19754e+14	21244

Table 8. Solution characteristics for different values of the capacity bounds factor.

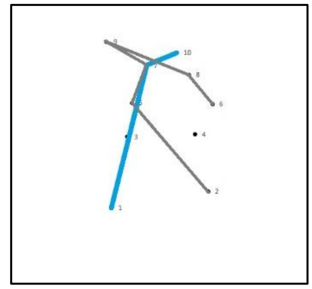
Row	υ_{ECap}	υ_{NCap}	Satisfied demand (%)	Profit value	CPU Time
1	0.5	0.5	11.47	4.15654e+13	2465
2	1	1	18.47	7.47316e+13	2709
3	2	2	26.36	1.10205e+14	2949



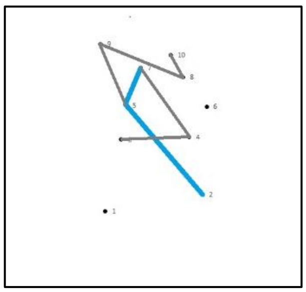
(a): 1st row of Table 6



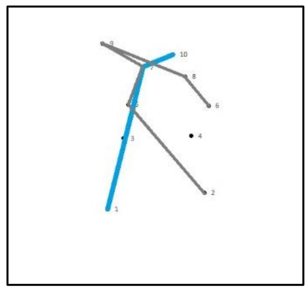
(b) 2nd row of Table 6



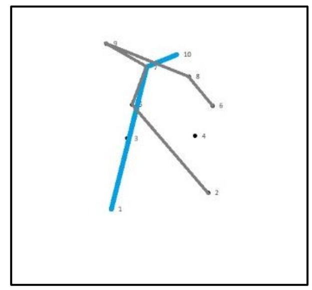
(c): 3rd row of Table 6



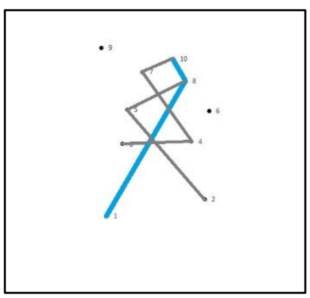
(d): 1st row of Table 7



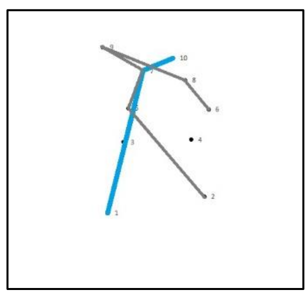
(e): 2nd row of Table 7



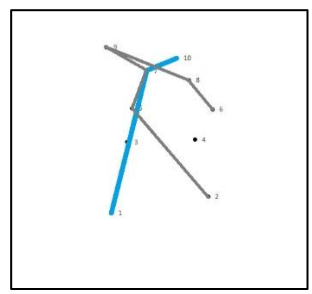
(f): 3rd row of Table 7



(g): 1st row of Table 8



(h): 2nd row of Table 8



(i): 3rd row of Table 8

Figure 2. Configuration of the networks obtained corresponding to the experiments of the mentioned row of Tables 6 to 8.

5. Conclusion

In this study, a novel mathematical programming model is presented for designing a rapid transit network based on a hub and spoke model under uncertainty, considering the hub and spoke nodes and edge capacity constraints. The network consists of stopovers in the hub and spoke alignments and both hub-level and spoke-level sub-networks are composed of multiple lines. The model decides about the location of the hub and spoke nodes, the allocation of the spoke nodes to the hub nodes, the selection of the hub and spoke edges, the determination of the hub and spoke lines, the determination of the percentage of satisfied OD demands, and the routing of satisfied demand flows through the lines. Capacity constraints are considered for the hub and spoke nodes and also the hub and spoke edges. The uncertainty is assumed for the demands and the transportation costs, represented by a finite set of scenarios. The goal is to maximize the total expected profit by considering transfer costs.

The model is evaluated on instances derived from the well-known AP dataset for hub problems, using the CPLEX solver. Comprehensive computational experiments suggest the effectiveness and efficiency of the proposed model in finding optimal or near optimal solution in a reasonable time and the relevance of considering a stochastic modeling framework for the problem. Some useful insight is also provided into the interactions among different aspects of the studied complex decision problem by measuring the sensitivity of the model to variations in some important parameters and analyzing the resulting hub networks under various parameter settings. It can be concluded that making different decisions to design a hub-based RTN and considering different aspects of the problem such as uncertainty and nodes and edges capacity constraints can lead to better solutions for real-world situations. For future research, developing efficient solution algorithms, either exact or meta-heuristic, instead of using commercial solvers, to be able to solve realistic-size instances of the problem is recommended. From the modeling point of view, even more, applicable models can be found, for example, by incorporating tactical decisions.

6. References

[1] Aykin, T. (1994), Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem, *European Journal of Operational Research*, 79 (3), 501-523.

[2] Basirati, M., Akbari Jokar, M. R. and Hassannayebi, E. (2020), Bi-objective optimization approaches to many-to-many hub location routing with distance balancing and hard time window, *Neural Computing and Applications*, 32 (17), 13267-13288.

[3] Birge, J. and FV, L. (2011), Introduction to stochastic programming, Springer, New York, Dordrecht, Heidelberg, London.

[4] Campbell, J. F., Ernst, A. T. and Krishnamoorthy, M. (2005), Hub Arc Location Problems: Part I—Introduction and Results, *Management Science*, 51 (10), 1540-1555.

[5] Canca, D., De-Los-Santos, A., Laporte, G. and Mesa, J. A. (2017), An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem, *Computers & Operations Research*, 78 1-14.

[6] Canca, D., De-Los-Santos, A., Laporte, G. and Mesa, J. A. (2019), Integrated Railway Rapid Transit Network Design and Line Planning problem with maximum profit, *Transportation Research Part E: Logistics and Transportation Review*, 127 1-30.

[7] Chen, K., Xin, X., Zhang, T. and Yang, Z. (2020), Multiport cooperative location model with a safe-corridors setting in West Africa, *International Journal of Logistics Research and Applications*, 23 (6), 580-601.

[8] Contreras, I., Fernández, E. and Marín, A. (2010), The Tree of Hubs Location Problem, *European Journal of Operational Research*, 202 (2), 390-400.

[9] Contreras, I., Cordeau, J.-F. and Laporte, G. (2011), Stochastic uncapacitated hub location, *European Journal of Operational Research*, 212 (3), 518-528.

[10] Contreras, I., Tanash, M. and Vidyarthi, N. (2013), The cycle hub location problem., *Technical Report CIRRELT 59.*,

[11] Contreras, I. and O'Kelly, M. (2019), Hub Location Problems. In: G. Laporte, S. Nickel, and F. Saldanha da Gama Eds, Location Science, Springer International Publishing: Cham, 327-363

[12] Correia, I., Nickel, S. and Saldanha-da-Gama, F. (2018), A stochastic multi-period capacitated multiple allocation hub location problem: Formulation and inequalities, *Omega*, 74 122-134.

[13] Danach, K., Gelareh, S. and Neamatian Monemi, R. (2019), The capacitated single-allocation p-hub location routing problem: a Lagrangian relaxation and a hyper-heuristic approach, *EURO Journal on Transportation and Logistics*, 8 (5), 597-631.

[14] Fallah-Tafti, M., Honarvar, M., Tavakkoli-Moghaddam, R. and Sadegheih, A. (2022), Mathematical modeling of a bi-objective hub location-routing problem for rapid transit networks, *RAIRO-Oper. Res.*, 56 (5), 3733-3763.

[15] Gelareh, S. and Nickel, S. (2011), Hub location problems in transportation networks, *Transportation Research Part E: Logistics and Transportation Review,* 47 (6), 1092-1111.
[16] Gutiérrez-Jarpa, G., Obreque, C., Laporte, G. and Marianov, V. (2013), Rapid transit network design for optimal cost and origin–destination demand capture, *Computers & Operations Research,* 40 (12), 3000-3009.
[17] Gutiérrez-Jarpa, G., Laporte, G., Marianov, V. and Moccia, L. (2017), Multi-objective rapid transit network design with modal competition: The case of Concepción, Chile, *Computers & Operations Research,* 78 27-43.
[18] Huang, D., Liu, Z., Fu, X. and Blythe, P. T. (2018), Multimodal transit network design in a hub-and-spoke network framework, *Transportmetrica A: Transport Science,* 14 (8), 706-735.
[19] Kartal, Z., Hasgul, S. and Ernst, A. T. (2017), Single allocation p-hub median location and routing problem with simultaneous pick-up and delivery, *Transportation Research Part E: Logistics and Transportation Review*, 108 141-159.
[20] Kaveh, F., Tavakkoli-Moghaddam, R., Triki, C., Rahimi, Y. and Jamili, A. (2019), A new bi-objective model of the urban public transportation hub network design under uncertainty, *Annals of Operations Research,*
[21] Kemmar, O., Bouamrane, K. and Gelareh, S. (2021), Hub location problem in round-trip service applications, *RAIRO-Oper. Res.*, 55 S2831-S2858.
[22] Labbé, M. and Yaman, H. (2008), Solving the hub location problem in a star–star network, *Networks,* 51 (1), 19-33.
[23] 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*, 136 (1), 49-63.
[24] Laporte, G., Marín, Á., Mesa, J. A. and Ortega, F. A. (2007), An Integrated Methodology for the Rapid Transit Network Design Problem. In: Algorithmic Methods for Railway Optimization, Springer: Berlin Heidelberg, 187-199.
[25] Laporte, G., Marín, A., Mesa, J. A. and Perea, F. (2011), Designing robust rapid transit networks with alternative routes, *Journal of Advanced Transportation*, 45 (1), 54-65.
[26] Laporte, G. and Mesa, J. A. ,(2019), The Design of Rapid Transit Networks. In: G. Laporte, S. Nickel, and F. Saldanha da Gama Eds, Location Science, Springer International Publishing: Cham, 687-703
[27] López-Ramos, F., Codina, E., Marín, Á. and Guarnaschelli, A. (2017), Integrated approach to network design and frequency setting problem in railway rapid transit systems, *Computers & Operations Research,* 80 128-146.
[28] Mahéo, A., Kilby, P. and Hentenryck, P. V. (2019), Benders Decomposition for the Design of a Hub and Shuttle Public Transit System, *Transportation Science,* 53 (1), 77-88.
[29] Mahmutoğulları, A. İ. and Kara, B. Y. (2015), Hub Location Problem with Allowed Routing between Nonhub Nodes, *Geographical Analysis,* 47 (4), 410-430.
[30] Marín, Á. and García-Ródenas, R. (2009), Location of infrastructure in urban railway networks, *Computers & Operations Research,* 36 (5), 1461-1477.
[31] Martins de Sá, E., Contreras, I. and Cordeau, J.-F. (2015), Exact and heuristic algorithms for the design of hub networks with multiple lines, *European Journal of Operational Research,* 246 (1), 186-198.
[32] Martins de Sá, E., Contreras, I., Cordeau, J.-F., de Camargo, R. S. and de Miranda, G. (2015), The Hub Line Location Problem, *Transportation Science,* 49 (3), 500-518.
[33] Oliveira, F. A., de Sá, E. M. and de Souza, S. R. (2022), Benders decomposition applied to profit maximizing hub location problem with incomplete hub network, *Computers & Operations Research,* 142 105715.
[34] Pourmohammadi, P., Tavakkoli-Moghaddam, R., Rahimi, Y. and Triki, C. (2021), Solving a hub location-routing problem with a queue system under social responsibility by a fuzzy meta-heuristic algorithm, *Annals of Operations Research,*

[35] Rahimi, Y., Torabi, S. A. and Tavakkoli-Moghaddam, R. (2019), A new robust-possibilistic reliable hub protection model with elastic demands and backup hubs under risk, *Engineering Applications of Artificial Intelligence,* 86 68-82.

[36] Rouzpeykar, Y., Soltani, R. and Kazemi, M. A. A. (2020), A Robust Optimization Model for the Hub Location and Revenue Management Problem Considering Uncertainties, *Iranian Journal of Operations Research,* 11 (1),

[37] Shushan, C., Qinghuai, L. and Zhong, S. (2019), Design of Urban Rail Transit Network Constrained by Urban Road Network, Trips and Land-Use Characteristics, *Sustainability,* 11

[38] Taherkhani, G., Alumur, S. A. and Hosseini, M. (2020), Benders Decomposition for the Profit Maximizing Capacitated Hub Location Problem with Multiple Demand Classes, *Transportation Science,* 54 (6), 1446-1470.

[39] Taherkhani, G., Alumur, S. A. and Hosseini, M. (2021), Robust Stochastic Models for Profit-Maximizing Hub Location Problems, *Transportation Science,* 55 (6), 1322-1350.

[40] Tavassoli, K. and Tamannaei, M. (2020), Hub network design for integrated Bike-and-Ride services: A competitive approach to reducing automobile dependence, *Journal of Cleaner Production*, 248 119247.

[41] Thomadsen, T. and Larsen, J. (2007), A hub location problem with fully interconnected backbone and access networks, *Computers & Operations Research,* 34 (8), 2520-2531.

[42] Tikani, H., Honarvar, M. and Mehrjerdi, Y. Z. (2018), Developing an integrated hub location and revenue management model considering multi-classes of customers in the airline industry, *Computational and Applied Mathematics,* 37 (3), 3334-3364.

[43] Tootooni, B., Sadegheih, A., Zare, H. K. and Vahdatzad, M. A. (2020), A novel type I and II fuzzy approach for solving single allocation ordered median hub location problem, 11 (2), 65-79.

[44] Verma, A., Kumari, A., Tahlyan, D. and Hosapujari, A. B. (2017), Development of hub and spoke model for improving operational efficiency of bus transit network of Bangalore city, *Case Studies on Transport Policy,* 5 (1), 71-79.

[45] Yaman, H., Kara, B. Y. and Tansel, B. Ç. (2007), The latest arrival hub location problem for cargo delivery systems with stopovers, *Transportation Research Part B: Methodological,* 41 (8), 906-919.