Transportation Research Part B 46 (2012) 529–543

Contents lists available at SciVerse ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Hub location under uncertainty

Sibel A. Alumur [a,*], Stefan Nickel [b,c], Francisco Saldanha-da-Gama [d]

[a] Department of Industrial Engineering, TOBB University of Economics and Technology, Ankara, Turkey
[b] Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[c] Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany
[d] DEIO-CIO, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

ARTICLE INFO

Article history:
Received 6 May 2011
Received in revised form 14 November 2011
Accepted 14 November 2011

Keywords:
Hub location
Uncertainty
Stochastic demands
Robust optimization

ABSTRACT

Hub location problems are network design problems which are solved as part of a strategic decision making process. In strategic planning, decisions may have a long lasting effect and the implementation may take considerable time. Moreover, input data is not precisely known in advance. Hence, decisions have to be made anticipating uncertainty. In this paper, we address several aspects concerning hub location problems under uncertainty. Two sources of uncertainty are considered: the set-up costs for the hubs and the demands to be transported between the nodes. Generic models are presented for single and multiple allocation versions of the problems. Firstly, the two sources of uncertainty are analyzed separately and afterwards a more comprehensive model is proposed considering all sources of uncertainty. Using a set of computational tests performed, we analyze the changes in the solutions driven by the different sources of uncertainty considered isolated and combined.

1. Introduction

In hub location problems, there is a given set of nodes with corresponding demands to be shipped among them. The problem consists of selecting a subset of the nodes to become hubs; i.e., nodes that consolidate and distribute the flow. One advantage for considering such type of facilities regards the potential saving in the costs due to economies of scale. A decision that is made together with the location of the hubs is the allocation of the non-hub nodes (spokes) to the selected hubs. The sub-network defined by the hubs and by the links between them is often called the hub network, whereas the sub-network defined by the spokes and by their links to the hubs is called the spoke network.

Hub location problems have applications in many areas such as public transportation systems, logistics distribution systems, and telecommunications.

Since the seminal papers by O'Kelly (1986, 1987) much work has been developed in this area. In particular, problems with different features and goals have been addressed. The *p*-hub median problem, and the capacitated and uncapacitated hub location problems are among the most frequently studied hub location problems in the literature. In the *p*-hub median problem, the objective is the minimization of the network operation costs (costs for routing the flow); in capacitated and uncapacitated hub location problems, on the other hand, set-up costs for the hubs are additionally considered in the objective function.

One important aspect in a hub location problem regards the allocation pattern of the spokes to the hubs. One among two possibilities is often considered: single allocation and multiple allocation. In the first case, each spoke is connected to a single hub. This is the case, for instance in Abdinnour-Helm (1998), Ernst and Krishnamoorthy (1996, 1999), Labbé and Yaman (2004), Labbé et al. (2005), Pirkul and Schilling (1998), and Skorin-Kapov and Skorin-Kapov (1994). In the second case, a spoke can send and receive flow from more than one hub. Some works considering this allocation pattern include Boland

* Corresponding author. Tel.: +90 312 2924136; fax: +90 312 2924091.
E-mail addresses: salumur@etu.edu.tr (S.A. Alumur), stefan.nickel@kit.edu (S. Nickel), fsgama@fc.ul.pt (F. Saldanha-da-Gama).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2011.11.006

et al. (2004), Contreras et al. (2012, in press), Contreras et al. (2009), Ebery et al. (2000), Ernst and Krishnamoorthy (1998a), Hamacher et al. (2004), Marin (2005), and Mayer and Wagner (2002). Some works address both situations as it is the case in the papers by Campbell (1994, 1996), Ernst and Krishnamoorthy (1998b), O'Kelly et al. (1996), and Skorin-Kapov et al. (1996). More recently, Yaman (2011) introduce the *r*-allocation *p*-hub median problem, in which more flexible allocation patterns of spokes to hubs are considered, thus, avoiding the classical single or multiple allocation schemes.

When some type of constraint exists in the flow that can be shipped through the network, the problem becomes capacitated. Depending on the problem, capacity constraints may refer to the nodes or to the edges. In the first case, we can still distinguish between a constraint on the total flow shipped through a hub (e.g. Campbell, 1994) or the non-processed flow entering the hubs (e.g. Ebery et al., 2000, Ernst and Krishnamoorthy, 1999). Some works combine capacity constraints both in the hubs and in the links such as Sasaki and Fukushima (2003).

One assumption that is often considered in the literature due to the large number of situations in which it applies is that the hub network is a complete graph, which means that all hubs are directly connected. Note, however, that some applications exist in which this is not the case (e.g. Alumur and Kara, 2009; Alumur et al., 2009; Calik et al., 2009; Campbell et al., 2005a,b; Contreras et al., 2009, 2010; Nickel et al., 2001; Yaman, 2008).

Another common assumption in hub location problems is that no direct shipment is possible between the spoke nodes. This is the case when all the flow must go through some type of processing such as in mail distribution systems. Nevertheless, in some applications, this assumption is relaxed (e.g. Aykin, 1994; Nickel et al., 2001; Yoon and Current, 2008).

Several mathematical programming based approaches and several heuristic procedures have been proposed for hub location problems. For a deeper overview on the models and solution techniques that have been developed for hub location problems the reader can refer to Alumur and Kara (2008) and to the references therein.

One aspect of major importance in many decision making problems regards the need to deal with uncertain data. Hub location problems are no exception. In fact, a hub location problem is a network design problem which is often solved as part of a strategic decision making process and thus, the solution may have a long lasting effect and its implementation may take considerable time. Moreover, often, such implementation must be finished before the network starts being operated. Accordingly, several parameters involved in hub location problems may not be known precisely when the network design decisions are made. This is typically the case with the set-up costs for the facilities and with the demands to be transported between the nodes.

Regarding the set-up costs, although they may be estimated before the decision is made, in practice, the actual set-up costs will vary due to many factors such as the price of the property (e.g. for land acquisition) and the price of the raw-materials (e.g. for constructing facilities).

As far as the demands are concerned, it can also be estimated in advance. However, the time elapsed between the moment the decision is made and the moment the network starts operating may make such information completely obsolete. For instance, demand may be stimulated or discouraged by the advent of the new facility. In any case, demands are hardly deterministic or known too much in advance.

When uncertainty is associated with some data involved in a decision making problem, we can distinguish between two situations: (i) no probability law can be associated with the uncertain parameters; (ii) uncertainty is described by some known probability distribution. In the first case, it is commonly assumed that the uncertain vector of parameters can take any value in some set (e.g. an interval) although no probability distribution function is known describing its behavior. This often happens with costs such as the set-up costs. In the second case, the vector of parameters is not known in advance but some probabilistic model has been identified which describes its behavior. In this case, we are facing a stochastic problem.

Independent from the existence of a probability law describing the uncertainty, we can make a further distinction depending on whether or not the possible realizations of the uncertain vector of parameters define a discrete set. When it does, each possible realization of the random vector is called a scenario. In such a case, it is usually assumed that a realization of the uncertain vector defines a value for each uncertain parameter; i.e., the uncertain parameters are correlated. At a first glance, this might seem an oversimplification of reality. However, this assumption is not very restrictive because, for example, in a situation with independent demands, one can consider the set of scenarios for each particular demand and combine all possibilities thus falling in the situation introduced above.

To the best of the authors' knowledge, only four papers exist in the literature addressing uncertainty issues in the context of hub location problems. Marianov and Serra (2003) consider a hub location problem in the context of airline transportation. Hubs are modeled as M/D/c queuing systems. A formula is derived for the probability of the number of customers in the system, which is later used to propose a capacity constraint. This constraint limits the probability of more than a certain number of airplanes in queue, to be smaller than or equal to a given value. The model is solved using a heuristic based on tabu search. The methodology is tested on the CAB data (O'Kelly, 1987) and also on randomly generated data. Yang (2009) addresses a problem also in the context of airline transportation. Stochasticity is associated with the demands and the problem is formulated as a two stage stochastic program. The deterministic equivalent problem is solved considering three scenarios. Direct shipment between spoke nodes is allowed. A case study is presented from the air freight market in Taiwan and China where only a single hub is located. Sim et al. (2009) consider a problem with normally distributed stochastic travel times. There is a constraint ensuring that the probability of the total travel time along a path being less than the given time bound is greater than a given value (chance constraints). Three scenarios with different coefficient of variation are analyzed. A heuristic algorithm is proposed and a set of computational tests are performed using the CAB and AP (Ernst and Krishnamoorthy,

1996) data sets. More recently, Contreras et al. (2011) studied stochastic uncapacitated multiple allocation hub location problems in which uncertainty is associated to demands and transportation costs. They showed that the problems with stochastic demands and stochastic dependent transportation costs have deterministic equivalents. They proposed an algorithm for the problem with stochastic independent transportation costs. They tested their algorithm on randomly generated instances and on the AP data set (Ernst and Krishnamoorthy, 1996).

Although addressing uncertainty issues in the context of hub location problems, the papers above do not give an insight on the impact of different sources of uncertainty on the solutions.

Considering a more general network structure, Lium et al. (2009) addresses demand stochasticity in service network design. The problem consists of planning the flow consolidation and shipment between differen origin/destination pairs. The goal is to select the services to include in the transportation plan, to build the service schedule and to determine how the flow is to be shipped through the network so that the total operational cost is minimized. Demand is assumed to be stochastic. The operational costs are assumed to be deterministic and no network design costs are involved. The experiments show that solutions based on stochastic approach can be structurally different from the deterministic counterparts and that consolidation and hub-and-spoke systems offer better solutions when there are uncertainties in the demand. The findings indicate that making strategic decisions, such as where to locate hubs, and more tactical/operational decisions, such as creating schedules, should preferably be a single decision process.

In this paper, we address hub location problems under uncertainty in the set-up costs and in the demands. We aim at presenting generic models capturing the uncertainty associated with the data. By performing computational analysis, we analyze the changes in the solutions with different sources of uncertainty. Our aim is to get an insight on how relevant a modeling framework comprising uncertainty aspects may be in hub location problems. Similar to the analysis in Lium et al. (2009), we study how solutions obtained under uncertainty differ from the solutions based on deterministic data.

We only consider discrete type of uncertainty where uncertainty is represented by a finite set of scenarios. When this is not the situation, one can resort to an approximate sampling technique (e.g. Birge and Louveaux, 1997; Kleywegt et al., 2001) and reduce the problem to the finite-scenario setting that we address.

The final modeling framework, which combines uncertainty in the set-up costs and in the demands, is built throughout the paper. We start by considering the simplest setting which refers to uncertainty only in the set-up costs. Then, we consider uncertainty only in the demands. Finally, we consider both sources of uncertainty in a single and more comprehensive modeling framework.

The remainder of this paper is organized as follows. In the next section, we present the basic setting upon which our modeling framework is built. We also present several underlying assumptions and most of the notation considered throughout the paper. In Section 3, we address hub location problems with uncertainty in the set-up costs, whereas, in the following section, we consider stochastic demands. In Section 5, we gather both sources of uncertainty in a single modeling framework. Throughout the paper, we present computational results that allow the reader to get a better understanding on how the solutions change with different features of uncertainty.

2. Basic setting

In this section, we introduce the basic setting for our analysis. We consider a simple base situation in order to get a more focused demonstration of the methodology we propose. Nevertheless, it should be noted that the contents of the following sections can be extended to more complex hub location models than those we address.

We consider uncapacitated hub location problems under the assumption that the hub level network is a complete graph. Additionally, we assume that there are no direct links between non-hub nodes; i.e., all traffic should be routed via at least one hub. Concerning the cost structure, we consider set-up costs for the hubs and flow shipment costs. Regarding the latter, we distinguish between collection, distribution, and transfer costs respectively for (i) the flow from spoke nodes to hubs, (ii) the flow between hubs, and (iii) the flow from hubs to spoke nodes. Accordingly, in terms of costs, we are assuming a structure that has been often considered in the literature (e.g. Contreras et al., 2009, 2010; Costa et al., 2008; Ernst and Krishnamoorthy, 1996, 1998a).

Hereafter, the following notation is considered.

$N = \{1,\ldots,n\}$	Set of nodes
f_k	Fixed set-up cost for installing a hub at node $k \in N$
d_{ij}	Distance from node $i \in N$ to node $j \in N$. It is assumed that $d_{ii} = 0$ ($i \in N$) and that the distances satisfy the triangle inequality
χ	Collection cost per unit of flow and per unit of distance
δ	Distribution cost per unit of flow and per unit of distance
α	Transfer cost, i.e., cost per unit of flow and per unit of distance between hubs. It is assumed that α is smaller than the collection and distribution costs
w_{ij}	Flow originated at node $i \in N$ that is destined to node $j \in N$
$O_i = \sum_{j\in N} w_{ij}$	Total flow originated at node $i \in N$
$D_i = \sum_{j\in N} w_{ji}$	Total flow destined to node $i \in N$

Several formulations have been proposed for the single allocation uncapacitated hub location problem as well as for the multiple allocation counterpart (Alumur and Kara, 2008).

Ernst and Krishnamoorthy (1996) proposed a set of 3-index flow variables for formulating hub location problems, which led to models that could better achieve a tradeoff between compactness and tightness. In this paper, we adopt such flow variables for both the single and the multiple allocation versions:

$y^i_{kl} =$	Amount of flow originated at node i that is routed via hubs k and l in this order $(i,k,l \in N)$

Considering these flow variables, the single allocation hub location problem (SAHLP) can be formulated as follows (Ernst and Krishnamoorthy, 1999):

$$\min \sum_{k\in N} f_k x_{kk} + \sum_{i\in N}\sum_{k\in N} \chi d_{ik} O_i x_{ik} + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y^i_{kl} + \sum_{i\in N}\sum_{k\in N} \delta d_{ki} D_i x_{ik} \tag{1}$$

$$\text{s.t.} \sum_{k\in N} x_{ik} = 1 \quad i \in N \tag{2}$$

$$x_{ik} \leqslant x_{kk} \quad i,k \in N \tag{3}$$

$$\sum_{l\in N} y^i_{kl} - \sum_{l\in N} y^i_{lk} = O_i x_{ik} - \sum_{j\in N} w_{ij} x_{jk} \quad i,k \in N \tag{4}$$

$$\sum_{l\in N, l\neq k} y^i_{kl} \leqslant O_i x_{ik} \quad i,k \in N \tag{5}$$

$$x_{ik} \in \{0,1\} \quad i,k \in N \tag{6}$$

$$y^i_{kl} \geqslant 0 \quad i,k,l \in N \tag{7}$$

In this formulation, x_{ik} is a binary variable which is equal to 1 or 0 depending on whether node i is allocated to hub k. $x_{kk} = 1$ simply indicates that node k is chosen to become a hub.

The objective function (1) represents the total cost to be minimized. Constraints (2) assure that each node is a hub or is allocated to a hub. Constraints (3) make sure that a non-hub node is only allocated to operating hubs. Constraints (4) are flow balance constraints. Constraints (5) are not in the model proposed by Ernst and Krishnamoorthy (1999). However, using the same arguments as pointed out by Correia et al. (2010) one realizes that these constraints are in fact necessary, if triangle inequality is not strictly satisfied. Accordingly, they are considered to assure the validity of the model independently from the structure of the data considered. Finally, constraints (6) and (7) are domain constraints.

Using the same 3-index flow variables introduced above, we can consider the following formulation, based on the formulation introduced in Ernst and Krishnamoorthy (1998a), for the multiple allocation hub location problem (MAHLP):

$$\min \sum_{k\in N} f_k x_k + \sum_{i\in N}\sum_{k\in N} \chi d_{ik} u_{ik} + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y^i_{kl} + \sum_{i\in N}\sum_{l\in N}\sum_{j\in N} \delta d_{lj} v^i_{lj} \tag{8}$$

$$\text{s.t.} \sum_{k\in N} u_{ik} = O_i \quad i \in N \tag{9}$$

$$\sum_{l\in N} v^i_{lj} = w_{ij} \quad i,j \in N \tag{10}$$

$$\sum_{l\in N} y^i_{kl} - \sum_{l\in N} y^i_{lk} = u_{ik} - \sum_{j\in N} v^i_{kj} \quad i,k \in N \tag{11}$$

$$\sum_{l\in N} y^i_{kl} \leqslant u_{ik}, \quad i,k \in N \tag{12}$$

$$u_{ik} \leqslant O_i x_k, \quad i,k \in N \tag{13}$$

$$v^i_{lj} \leqslant w_{ij} x_l, \quad i,j,l \in N \tag{14}$$

$$x_k \in \{0,1\} \quad k \in N \tag{15}$$

$$y^i_{kl}, u_{ik}, v^i_{jl} \geqslant 0 \quad i,j,k,l \in N \tag{16}$$

In the above model, u_{ik} represents the total flow that is sent from node i directly to hub k; v^i_{lj} is the flow originated at node i that flows from hub l to node j; x_k is a binary variable that is equal to 1 if node k is selected to be a hub and 0 otherwise.

The objective function (8) represents the total costs for building and operating the network. Constraints (9) (constraints (10)) assure that all flow is collected at (distributed from) the hubs. Constraints (11) are flow divergence constraints. Constraints (12) and (14) are consistency constraints that make the definition of the variables meaningful. Finally, constraints (15) and (16) are domain constraints.

3. Uncertainty in the set-up costs

As it was mentioned in the introductory section, uncertainty in the set-up costs may be driven by certain factors such as the cost of land acquisition or construction costs (which, in turn, may depend on other factors) and thus it can hardly be described by some probability law. Accordingly, it is assumed that no probabilistic information can be associated with these unknown parameters. Nevertheless, the uncertainty associated with the set-up costs is assumed to be fully described by a set of scenarios.

Without a probabilistic measure for the uncertainty, a possibility is to use a robustness measure in order to evaluate the performance of the system. One common such measure is the regret of a solution (the reader can refer to Kouvelis and Yu (1997) and Snyder (2006) for further details on robustness measures) which is the difference between the cost of a solution for some scenario and the optimal cost that can be achieved under that scenario. We propose the use of a minmax regret formulation as a mean to deal with the uncertainty in the set-up costs. Accordingly, we propose a formulation aiming at minimizing the worst case (maximum) regret over all scenarios. In addition to the parameters introduced in the previous section, we now define:

S_f	Set of scenarios for the uncertain set-up costs
f_k^s	Set-up cost for establishing a hub at node $k \in N$ under scenario $s \in S_f$

3.1. Single allocation

Consider that a single allocation pattern is assumed. For each scenario $s \in S_f$ we can formulate the corresponding problem:

$$min \sum_{k \in N} f_k^s x_{kk} + \sum_{i \in N} \sum_{k \in N} \chi d_{ik} O_i x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} y_{kl}^i + \sum_{i \in N} \sum_{k \in N} \delta d_{ki} D_i x_{ik} \tag{17}$$
$$s.t.\ (2)-(7).$$

We denote by Z_s^* the optimal value of the above problem ($s \in S_f$).

As we do not know the exact scenario that will occur, we can consider a minmax regret model namely

$$min \max_{s \in S_f} R_s \tag{18}$$
$$s.t.\ (2)-(7),$$
$$R_s = \left(\sum_{k \in N} f_k^s x_{kk} + \sum_{i \in N} \sum_{k \in N} \chi d_{ik} O_i x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} y_{kl}^i + \sum_{i \in N} \sum_{k \in N} \delta d_{ki} D_i x_{ik} \right) - Z_s^* \quad s \in S_f \tag{19}$$

The above model can be linearized straightforwardly by defining a new variable R and setting it greater or equal to all R_s, $s \in S_f$.

Note that none of the decision variables, other than regret, has a scenario index. Thus, this formulation is rather simple. Nevertheless, we decided to consider it as part of our analysis so that a clear perception can be obtained regarding the impact from considering this source of uncertainty in the solutions.

3.2. Multiple allocation

Following the same reasoning presented above, we can address the multiple allocation case. Accordingly, for each scenario $s \in S_f$, we consider the following problem:

$$min \sum_{k \in N} f_k^s x_k + \sum_{i \in N} \sum_{k \in N} \chi d_{ik} u_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} y_{kl}^i + \sum_{i \in N} \sum_{l \in N} \sum_{j \in N} \delta d_{lj} v_{lj}^i \tag{20}$$
$$s.t.\ (9)-(16).$$

As before, we denote by Z_s^* the optimal value of the above problem ($s \in S_f$). The corresponding minmax regret model becomes:

$$min \max_{s \in S_f} R_s \tag{21}$$
$$s.t.\ (9)-(16),$$
$$R_s = \left(\sum_{k \in N} f_k^s x_k + \sum_{i \in N} \sum_{k \in N} \chi d_{ik} u_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} y_{kl}^i + \sum_{i \in N} \sum_{l \in N} \sum_{j \in N} \delta d_{lj} v_{lj}^i \right) - Z_s^* \quad s \in S_f \tag{22}$$

As we noted for the single allocation case, the above model is straightforwardly linearizable. Again, none of the decision variables, other than regret, has a scenario index.

3.3. Computational analysis

In this section, we present computational analyses with the single and multiple allocation minmax regret hub location models, to assess the effects of uncertainty in the set-up costs on the resulting solutions. For this analysis, we use the

Table 1
Results with the single allocation minmax regret model with uncertain set-up costs.

	$\alpha = 0.2$			$\alpha = 0.4$			$\alpha = 0.6$			$\alpha = 0.8$		
	Set-up costs	Trans. costs	Hub locations	Set-up costs	Trans. costs	Hub locations	Set-up costs	Trans. costs	Hub locations	Set-up costs	Trans. costs	Hub locations
Base case	427.64	541.20	4,7,12,17,24	338.98	794.57	1,4,12,18	338.98	941.12	1,4,12,18	253.27	1158.84	2,4,12
Scenario1, S_f^1	354.94	576.30	9,11,12,17,24	354.94	744.70	9,11,12,17,24	354.94	912.35	9,11,12,17,24	231.68	1188.79	5,8,17
Scenario2, S_f^2	361.78	553.86	4,12,14,16,17	352.05	731.74	4,12,14,16,18	352.05	897.32	4,12,14,16,18	179.05	1187.22	5,8,18
Scenario3, S_f^3	326.56	621.82	6,11,18,22,24	322.88	789.11	6,13,18,22,24	322.88	949.22	6,13,18,22,24	213.49	1190.04	8,21,25
Scenario4, S_f^4	359.31	564.57	4,12,13,17,24	262.04	813.09	1,2,4,12	262.04	952.47	1,2,4,12	262.04	1091.84	1,2,4,12
Scenario5, S_f^5	345.81	572.44	5,7,12,14,17	344.43	738.84	6,12,14,17,21	344.43	900.47	6,12,14,17,21	215.92	1166.32	12,21,25
Minmax regret	–	576.30	9,11,12,17,24	–	744.70	9,11,12,17,24	–	982.29	5,12,18,24	–	1188.55	2,5,8

Table 2
Results with the multiple allocation minmax regret model with uncertain set-up costs.

	$\alpha = 0.2$			$\alpha = 0.4$			$\alpha = 0.6$			$\alpha = 0.8$		
	Set-up costs	Trans. costs	Hub locations	Set-up costs	Trans. costs	Hub locations	Set-up costs	Trans. costs	Hub locations	Set-up costs	Trans. costs	Hub locations
Base case	427.64	534.70	4,7,12,17,24	336.22	760.96	4,12,18,24	250.12	953.66	12,18,21	250.12	1021.49	12,18,21
Scenario 1, S_f^1	354.94	575.56	9,11,12,17,24	354.94	718.46	9,11,12,17,24	354.94	835.87	9,11,12,17,24	268.78	1014.72	9,16,17,19
Scenario 2, S_f^2	361.78	544.85	4,12,14,16,17	352.05	699.86	4,12,14,16,18	289.73	876.07	4,12,16,18	289.73	959.32	4,12,16,18
Scenario 3, S_f^3	326.56	617.56	6,11,18,22,24	326.56	756.58	6,11,18,22,24	250.03	945.80	6,13,18,22	209.47	1046.53	18,21,22
Scenario 4, S_f^4	359.31	552.48	4,12,13,17,24	263.94	779.27	2,4,12,24	258.23	886.78	2,4,12,13	199.24	1027.68	2,4,12
Scenario 5, S_f^5	344.43	567.41	6,12,14,17,21	344.43	703.59	6,12,14,17,21	344.43	826.42	6,12,14,17,21	217.84	1021.91	12,17,21
Minmax regret	–	575.56	9,11,12,17,24	–	760.96	4,12,18,24	–	874.47	4,12,18,24	–	1026.80	2,12,21

well-known CAB data set. The CAB data set is initially introduced by O'Kelly (1987) and presented in the OR-Library (2011). The data contains distances and demands based on airline passenger interactions between 25 US cities. The demands are scaled by dividing with the total demand so that the total demand is equal to 1.

For the CAB data set, the collection and distribution costs per unit are taken equal to one; i.e., $\chi = \delta = 1$. For the value of α, on the other hand, we let $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$, as it is customarily done in the literature.

We assume that relatively larger cities have larger set-up costs. Thus, for each city in the CAB data set, we correlate set-up costs with the amount of total demand originated at that city. Considering that the variation in the set-up costs should not be too much, we take the fixed set-up costs randomly from the interval $[10\log O_k, 20\log O_k]$ for each potential hub location $k \in N$. For example for node 1, f_1 is taken randomly from the interval [53.9, 107.7], for node 2 from [51.6, 103.1], and for node 3 from the interval [57.1, 114.3].

We generated five different scenarios with uncertain set-up costs. In each scenario $s \in S_f$ and for each $k \in N$, we let f_k^s to take a random value from the interval $[10\log O_k, 20\log O_k]$.

We solved the single and multiple allocation versions of the problems separately under each scenario and also with the minmax regret models. We compare these solutions with the solution obtained when the set-up costs are set to the mean values of the intervals; i.e., when $f_k = 15 \log O_k$ for all $k \in N$. We refer to this solution as the base case solution. We present the results in Table 1 for the single allocation and in Table 2 for the multiple allocation version of the problem.

All the instances are solved to optimality with CPLEX version 11.2 on a server with 2.6 GHz AMD Opteron 252 processor and 2 GB of RAM. We do not report the CPU times separately for each instance since all the instances are solved within a few seconds.

In Tables 1 and 2, for each α value, the ''Set-up costs'' columns list the value of the total set-up costs and ''Trans. costs'' columns list the total transportation costs in the optimal solution of the corresponding instances. ''Hub locations'' columns present the optimal hub locations. Note that in minmax regret solutions total set-up costs are different under each scenario, hence, the value of the set-up costs are left empty.

Observe from Tables 1 and 2 that, for each α value, the optimal hub locations are different in each of the scenarios. Moreover, the optimal numbers of established hubs are not always the same. Except at few instances, the optimal solutions of the minmax regret models are different than the solutions obtained under different scenarios.

The magnitude of the set-up costs in the objective function value changes with varying α values. When alpha value is set to 0.8, magnitude of the transportation costs are highest in the objective function value compared to the magnitude of the set-up costs. Even though the magnitude of the set-up costs are low when $\alpha = 0.8$, optimal hub locations are different under each of the scenarios and also with the minmax regret model. This proves that the optimal solutions are sensitive to the values of the set-up costs on the CAB data set.

When we look at the set-up costs of the optimal hub facilities, we do not observe any relationship between being chosen as a hub and the particular set-up cost of a node under each scenario. That is, the model is not simply choosing hubs at cities with lower set-up costs. For example, node 12 is the third node generating the highest amount of flow in the CAB data set, and thus, it is among the most expensive locations under all scenarios (always within the six most expensive locations). However, note that node 12 is selected as a hub in more than half of the optimal solutions. Other factors such as the demand and geographical location are also important for selecting a node as a hub. On the other hand, even though some particular hub locations are common at various instances in Tables 1 and 2, the resulting solutions are almost always different under different scenarios.

We may conclude that since optimal solutions are sensitive to the set-up costs, with high amount of uncertainty in the set-up costs, it is better to adopt the solution obtained with the minmax regret model instead of using an estimate for the set-up costs, or adopting the solution of a particular scenario.

4. Uncertainty in the demands

In this section, we consider a stochastic programming approach for hub location problems with uncertain demands. As in the previous section, we address both the single allocation and the multiple allocation cases.

The goal is to make network design decisions (namely the nodes that should become hubs and the allocation of the spoke nodes to the selected hubs) before knowing the demands. Although flows on the network are only revealed after knowing the demands, it is desirable to take such uncertainty into account when designing the network. This is what we propose below.

Independently from the allocation pattern, we assume that the uncertainty associated with demands can be described by a finite set of scenarios each of which having a probability that is assumed to be known. We assume that each scenario fully establishes a value for all the demands; i.e., we assume that demands are correlated. As mentioned before, this assumption is not very restrictive because in a situation with independent demands, we can consider the set of scenarios for each particular demand and then combine all possibilities thus resulting in the situation that we are addressing now.

As in Section 3, we need to introduce further notation:

S_w	Finite set of scenarios for the uncertain demands
p_s	Probability that scenario $s \in S_w$ occurs. We assume that $\sum_{s\in S_w} p_s = 1$
w_{ij}^s	Flow originated at node $i \in N$ and destined to node $j \in N$ under scenario $s \in S_w$
$O_i^s = \sum_{j\in N} w_{ij}^s$	Total flow originated at node $i \in N$ under scenario $s \in S_w$
$D_i^s = \sum_{j\in N} w_{ji}^s$	Total flow destined to node $i \in N$ under scenario $s \in S_w$

Denote by ξ the random variable which represents the scenario that occurs. Accordingly, $\xi \in S_w$ and $p_s = P[\xi = s]$, $s \in S_w$. Furthermore, $w_{ij}(\xi)$, $O_i(\xi)$ and $D_i(\xi)$ denote the random variables representing, respectively, the flow to be shipped from $i \in N$ to $j \in N$, the total flow originated at $i \in N$, and the total flow destined to $i \in N$.

4.1. Single allocation

Considering a single allocation pattern, as well as the decision variables introduced in Section 2, we can formulate the SAHLP with stochastic demands as follows:

$$\begin{aligned} min\ & \sum_{k\in N} f_k x_{kk} + \mathcal{Q}(\mathbf{x}) \\ s.t.\ & (2), (3), (6). \end{aligned} \tag{23}$$

where $\mathcal{Q}(\mathbf{x}) = E_\xi[Q(\mathbf{x},\xi)]$ and

$$Q(\mathbf{x},\xi) = min \sum_{i\in N}\sum_{k\in N} \chi d_{ik} O_i(\xi) x_{ik} + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y_{kl}^i + \sum_{i\in N}\sum_{k\in N} \delta d_{ki} D_i(\xi) x_{ik} \tag{24}$$

$$s.t. \sum_{l\in N} y_{kl}^i - \sum_{l\in N} y_{lk}^i = O_i(\xi) x_{ik} - \sum_{j\in N} w_{ij}(\xi) x_{jk} \quad i,k \in N \tag{25}$$

$$\sum_{l\in N, l\neq k} y_{kl}^i \leqslant O_i(\xi) x_{ik} \quad i,k \in N \tag{26}$$

$$y_{kl}^i \geqslant 0 \quad i,k,l \in N \tag{27}$$

The first stage problem determines the location and allocation decisions. In the second stage problem, the flows are consolidated and routed through the network. The objective function (23) evaluates the total set-up cost for the hubs plus the expected cost for consolidating and routing the flow.

In the second stage problem, the variables x_{ik} are fixed and, as introduced above, for all $i, j \in N$, $w_{ij}(\xi)$, $O_i(\xi)$, and $D_i(\xi)$ are random variables. The objective function (24) evaluates the total cost for consolidating and routing the flow. Constraints (25) and (26) assure correct flow consolidation and routing.

We are assuming that the demands are stochastic. Accordingly, the variables y^i_{kl} represent, in fact, a recourse decision because the exact way of shipping the flows through the network can only be defined after the demands are known. In addition, the coefficients associated with the recourse decision variables y^i_{kl} are deterministic. Accordingly, we are facing a problem with fixed recourse.

Due to the fact that no capacity constraints are being considered, once a feasible network design is obtained (i.e., the hubs and the allocation of spoke nodes to hubs) there is always a feasible solution in terms of flow consolidation and distribution. Accordingly, we are facing a problem with complete recourse. In fact, for each first-stage feasible solution (i.e. feasible network configuration) there is at least one second stage feasible solution (flow consolidation and distribution).

Due to the fact that the flow variables represent, in fact, a recourse decision, for each scenario $s \in S_w$, values must be found for all the flow variables; that is, we need to find $y^i_{kl}(s)$, which represents the best decision under scenario $s \in S_w$ for routing the flow originated at $i \in N$ that is sent through hubs $k \in N$ and $l \in N$. This aspect supports the redefinition of the recourse decision variables by considering a scenario index. In particular, we can consider

$y^{is}_{kl} =$	Amount of flow originated at node i that is routed between hubs k and l in this order under scenario s ($i,k,l \in N$, $s \in S_w$)

We can now write the so-called extensive form of the deterministic equivalent of SAHLP with stochastic demands:

$$\min \sum_{k \in N} f_k x_{kk} + \sum_{s \in S_w} p_s \left(\sum_{i \in N} \sum_{k \in N} \chi d_{ik} O^s_i x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} y^{is}_{kl} + \sum_{i \in N} \sum_{k \in N} \delta d_{ki} D^s_i x_{ik} \right) \tag{28}$$

$$\text{s.t. } (2), (3), (6),$$

$$\sum_{l \in N} y^{is}_{kl} - \sum_{l \in N} y^{is}_{lk} = O^s_i x_{ik} - \sum_{j \in N} w^s_{ij} x_{jk} \quad i, k \in N,\ s \in S_w \tag{29}$$

$$\sum_{l \in N, l \neq k} y^{is}_{kl} \leqslant O^s_i x_{ik} \quad i, k \in N,\ s \in S_w \tag{30}$$

$$y^{is}_{kl} \geqslant 0 \quad i, k, l \in N, \quad s \in S_w \tag{31}$$

Note that the first stage decision variables are common to all scenarios. Thus, the non-anticipativity constraints are implicit in the above model.

4.2. Multiple allocation

As in the single allocation situation, the goal is to make a decision about the network design before knowing the demands. Again, the flows on the network will only be disclosed sometime in the future after knowing the demands.

Considering all the notation already introduced as well as the considerations made in the beginning of Section 4.1, we can formulate the problem as follows:

$$\min \sum_{k \in N} f_k x_k + \mathcal{Q}(\mathbf{x}) \tag{32}$$

$$\text{s.t. } (15).$$

where $\mathcal{Q}(\mathbf{x}) = E[Q(\mathbf{x}, \xi)]$ and

$$Q(\mathbf{x}, \xi) = \min \sum_{i \in N} \sum_{k \in N} \chi d_{ik} u_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha d_{kl} y^i_{kl} + \sum_{i \in N} \sum_{l \in N} \sum_{j \in N} \delta d_{lj} v^i_{lj} \tag{33}$$

$$\text{s.t. } \sum_{k \in N} u_{ik} = O_i(\xi) \quad i \in N \tag{34}$$

$$\sum_{l \in N} v^i_{lj} = w_{ij}(\xi) \quad i, j \in N \tag{35}$$

$$\sum_{l \in N} y^i_{kl} - \sum_{l \in N} y^i_{lk} = u_{ik} - \sum_{j \in N} v^i_{kj} \quad i, k \in N \tag{36}$$

$$\sum_{l \in N} y^i_{kl} \leqslant u_{ik}, \quad i, k \in N \tag{37}$$

$$u_{ik} \leqslant O_i(\xi) x_k, \quad i, k \in N \tag{38}$$

$$v^i_{lj} \leqslant w_{ij}(\xi) x_l, \quad i, j, l \in N \tag{39}$$

$$y^i_{kl}, u_{ik}, v^i_{jl} \geqslant 0 \quad i, j, k, l \in N \tag{40}$$

The first stage problem determines the location decisions. In the second stage problem, the flows are consolidated and routed through the network. The objective function (32) in the first stage problem, evaluates the total set-up cost for the hubs plus the expected cost for consolidating and routing the flow.

In the second stage problem, the variables x_k are fixed and for all $i, j \in N$, $w_{ij}(\xi)$, $O_i(\xi)$, and $D_i(\xi)$ are random variables. The objective function (33) evaluates the total cost for consolidating and routing the flow. Constraints (34) assure that all flow is collected whereas (35) assure that all flow is distributed. Constraints (36) and (37) assure a correct flow consolidation and routing. Constraints (38) and (39) assure the consistency between the collection and distribution flows and the actual hubs. Finally, (40) are the domain constraints.

A decision about the flows can only be made after the demands occur. Accordingly, $y^i_{kl},\ u_{ik}$ and v^i_{jl} $(i,j,k,l \in N)$ represent a recourse decision.

As it was done for the single allocation situation, we can go further in terms of modeling the problem.

Due to the fact that the flow variables represent, in fact, a recourse decision, for each scenario $s \in S_w$ values must be found for all the flow variables; that is, we need to find $y^i_{kl}(s),\ u_{ik}(s)$ and $v^i_{lj}(s)$ for all $i, j, k, l \in N$, which, together, represent the best decision under scenario $s \in S_w$ for routing the flow. This aspect supports the redefinition of the recourse decision variables by considering a scenario index. In particular, we can consider

$y^{is}_{kl} =$	Amount of flow originated at node i that is routed between hubs k and l in this order under scenario s $(i,k,l \in N,$ $s \in S_w)$
$u^s_{ik} =$	Flow sent from node i directly to hub k under scenario s $(i,k \in N,\ s \in S_w)$
$v^{is}_{lj} =$	Flow originated at i that arrives at node j via hub l under scenario s $(i,j,l \in N,\ s \in S_w)$

We can now write the so-called extensive form of the deterministic equivalent of MAHLP with stochastic demands:

$$min \sum_{k\in N} f_k x_{kk} + \sum_{s\in S_w} p_s \left(\sum_{i\in N}\sum_{k\in N} \chi d_{ik} u^s_{ik} + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y^{is}_{kl} + \sum_{i\in N}\sum_{l\in N}\sum_{j\in N} \delta d_{lj} v^{is}_{lj} \right) \tag{41}$$

$s.t.$ (15),

$$\sum_{k\in N} u^s_{ik} = O^s_i \quad i \in N,\ s \in S_w \tag{42}$$

$$\sum_{l\in N} v^{is}_{lj} = w^s_{ij} \quad i,j \in N,\ s \in S_w \tag{43}$$

$$\sum_{l\in N} y^{is}_{kl} - \sum_{l\in N} y^{is}_{lk} = u^s_{ik} - \sum_{j\in N} v^{is}_{kj} \quad i,k \in N,\ s \in S_w \tag{44}$$

$$\sum_{l\in N} y^{is}_{kl} \leqslant u^s_{ik}, \quad i,k \in N,\ s \in S_w \tag{45}$$

$$u^s_{ik} \leqslant O^s_i x_k, \quad i,k \in N,\ s \in S_w \tag{46}$$

$$v^{is}_{lj} \leqslant w^s_{ij} x_l, \quad i,j,l \in N,\ s \in S_w \tag{47}$$

$$y^{is}_{kl}, u^s_{ik}, v^{is}_{jl} \geqslant 0 \quad i,j,k,l \in N,\ s \in S_w \tag{48}$$

In the next section, we present computational analysis with the deterministic equivalents of the single and multiple allocation hub location problems with stochastic demands.

4.3. Computational analysis

We test our stochastic single and multiple allocation hub location models with uncertain demands on the CAB data set. We use the same parameter setting for this data set as described in Section 3.3 and we use the set-up costs corresponding to the base case, where f_k = 15 logO_k for all $k \in N$.

For the stochastic models, we additionally require a different set of scenarios for the uncertain demands. For each node pair $i, j \in N$, we assume that the demands are realized from the interval $[0.01\bar{w}_{ij}, 10\bar{w}_{ij}]$, where $\bar{w}_{ij}$ are the demands given in the CAB data set. The reasoning behind the choice of this interval has to do with the need to obtain data with some relevant degree of variability, which is the minimum ingredient required to make a stochastic approach meaningful. By considering the above interval, the possible values for the demand range from one hundredth of $\bar{w}_{ij}$ and ten times this value. Such interval seemed to us as a reasonable choice for our purposes.

We generated five different scenarios. In order to reduce the symmetry around the mean in the generated scenarios, we split the interval into two parts: $[0.01\bar{w}_{ij}, 5\bar{w}_{ij}]$ and $[5\bar{w}_{ij}, 10\bar{w}_{ij}]$. In each scenario, for each $i, j \in N$ pair, the demand takes a random value from the first half of the interval with probability 2/3, and from the second half of the interval with probability 1/3. In each scenario, the demands are then scaled so that the total demand is always equal to 1. We took the probability of occurrence of each of the scenarios to be 1/5.

Table 3
Results with the stochastic single allocation model with uncertain demands.

	$\alpha = 0.2$		$\alpha = 0.4$		$\alpha = 0.6$		$\alpha = 0.8$	
	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.
Base case: $w_{ij} = \bar{w}_{ij}$	968.83	4,7,12,17,24	1133.55	1,4,12,18	1280.10	1,4,12,18	1412.10	2,4,12
Scenario 1, S_w^1	949.48	4,7,12,14,17	1113.41	4,12,17,24	1266.56	4,12,18,24	1382.76	2,5,8
Scenario 2, S_w^2	932.18	4,12,17,24	1088.21	4,12,17,24	1238.86	4,12,18,24	1357.30	2,5,12
Scenario 3, S_w^3	929.98	4,12,17,24	1085.12	4,12,17,24	1235.57	4,12,17,24	1354.48	2,4,12
Scenario 4, S_w^4	931.53	4,7,12,14,17	1092.75	4,12,18,24	1243.45	4,12,14,18	1366.04	2,4,12
Scenario 5, S_w^5	939.64	4,7,12,14,17	1102.07	4,12,14,17	1254.49	4,12,14,18	1373.91	2,5,12
Stochastic solution	937.74	4,12,17,24	1096.44	4,12,17,24	1248.54	4,12,14,18	1367.55	2,5,12
Added value	4.08		15.29		2.74		1.87	

Table 4
Results with the stochastic multiple allocation model with uncertain demands.

	$\alpha = 0.2$		$\alpha = 0.4$		$\alpha = 0.6$		$\alpha = 0.8$	
	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.
Base case: $w_{ij} = \bar{w}_{ij}$	962.34	4,7,12,17,24	1097.18	4,12,18,24	1203.77	12,18,21	1271.61	12,18,21
Scenario 1, S_w^1	942.31	4,12,17,24	1075.42	4,12,18,24	1184.38	4,12,18,24	1244.33	12,18,21
Scenario 2, S_w^2	923.90	4,12,17,24	1055.52	4,12,17,24	1163.68	4,12,18,24	1226.99	12,18,21
Scenario 3, S_w^3	920.87	4,12,17,24	1050.64	4,12,17,24	1159.22	4,12,18	1224.37	4,12,18
Scenario 4, S_w^4	924.06	4,12,17,24	1057.41	4,12,18,24	1168.11	4,12,18,24	1235.09	4,12,18
Scenario 5, S_w^5	931.42	4,12,17,24	1064.84	4,12,17,24	1174.23	4,12,18,24	1236.21	12,18,21
Stochastic solution	928.51	4,12,17,24	1061.18	4,12,17,24	1170.53	4,12,18,24	1234.77	12,18,21
Added value	7.74		0.76		7.31		0	

We then solved the deterministic equivalent of the single and multiple allocation hub location models with stochastic demands to optimality using CPLEX version 11.2 on the same computer with the properties given in Section 3.3. The results are presented in Tables 3 and 4.

Tables 3 and 4 list the optimal objective function values and the optimal hub locations separately under each of the scenarios and with the stochastic models. The last rows denoted by "Added value" present the added value of using the stochastic solution. All the instances listed in these tables are solved to optimality. The highest CPU time that we obtained among these instances is 10.6 minutes when $\alpha = 0.8$ with the stochastic multiple allocation model.

In the CAB data set, a high percentage of the total demand is generated only by a few nodes. Even though the interval for generating the demand values can be considered to be large, some nodes are still generating relatively higher percentage of the demand. For example, nodes 4 and 17 are the two cities generating the highest amount of flow in the CAB data set and also in three of the randomly generated scenarios (S_w^3, S_w^4, and S_w^5). In the remaining two scenarios (S_w^1 and S_w^2), these two cities are among the first four cities generating the highest amount of flow. Thus, it is not surprising that there is usually a hub located at nodes 4 and 17 in most of the instances. Because of these reasons, a few hub locations tend to stay optimal even with a high variability in the demands. We believe such an outcome is realistic, since in many cases demand is concentrated at some locations in spite of uncertainty.

Even though some hub locations tend to stay optimal, there are some instances in which the locations of some hub nodes differ from the base case solution. For example, with the single allocation problem when $\alpha = 0.4$, the hubs are located at nodes 1, 4, 12, and 18 in the base case. On the other hand, under three of the scenarios and with the stochastic model, the hubs are located at nodes 4, 12, 17, and 24. Although hubs 4 and 12 are common, with variation in the demand it is more advantageous to locate remaining hubs at nodes 17 and 24 compared to locating hubs at nodes 1 and 18. Similar situations are observed with some other values of α with the single and multiple allocation problems, where the locations of the hubs in the base case are different from the stochastic solutions.

There are some instances where the optimal number of hubs to be located in the stochastic solution is different from the base case solution. For example, in the single and multiple allocation problems when $\alpha = 0.2$, the total number of hubs to be located in the stochastic solution is one less than the base case solution. On the other hand, when $\alpha = 0.6$ in the multiple allocation problem, one more hub is required in the optimal stochastic solution. These results show that although total demand is the same, the optimal number of hubs required to provide service can differ under uncertainty in the demand.

To measure the improvement in the objective function value by using the stochastic solution compared to using the base case solution, we compute the added value of using the stochastic solution. This value is calculated as the difference between (i) the objective value in the stochastic problem by fixing the optimal base case solution and (ii) the optimal objective value of the stochastic problem. In order to determine the term mentioned in (i) we fix the locations of the hub nodes obtained in the base case solution and solve the deterministic model under different scenarios. We then calculate the expected objective function value considering the objective function values of the scenarios under the base case solution. The last rows in Tables 3 and 4 present the added value of using the stochastic solutions at different instances.

The added value of using the stochastic solutions ranges from 0 to 15.29 units. (The added value of 0 corresponds to the instance with the multiple allocation problem when $\alpha = 0.8$. Note that at this specific instance, the stochastic solution is exactly the same as the base case solution.) When these values are represented as a percentage of the total cost of the stochastic solution, they range from 0% to 1.39%. The added value of using the stochastic solution may not be considered to be high in the instances presented in Tables 3 and 4. This is due to the assumption that the total demand stays the same (always equal to 1). In real-life scenarios, we expect total demand to vary as well. The reason for not varying total demand in our scenarios is to isolate the effects of uncertainty on the individual demands. In general, we expect the added value to be higher when there is variation in the total demand.

The results from the analysis with the stochastic models show that if the decision maker does not consider uncertainty in the demands and adopts the base case solution (i.e., ignores the possibility of using a stochastic programming modeling framework), the resulting hub locations may not be optimal and total costs may increase up to 1.39% even if there is no variation in the total demand.

5. Different sources of uncertainty

In this section, we gather two types of uncertainty considered in the Sections 3 and 4 in a single modeling framework. Again, we distinguish between the single allocation and multiple allocation versions.

5.1. Single allocation

Considering all the information provided in Sub Sections 3.1 and 4.1, for each scenario $s' \in S_f$ we can consider the following stochastic single allocation hub location problem:

$$\begin{aligned} &\min \sum_{k\in N} f_k^{s'} x_{kk} + \sum_{s\in S_w} p_s \left(\sum_{i\in N}\sum_{k\in N} \chi d_{ik} O_i^s x_{ik} + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y_{kl}^{is} + \sum_{i\in N}\sum_{k\in N} \delta d_{ki} D_i^s x_{ik} \right) \quad (49)\\ &\quad s.t.\ (2), (3), (6), (29) - (31). \end{aligned}$$

Denote by $Z_{s'}^*$ the optimal value of the above problem. We can now propose a robust-stochastic model for the problem:

$$\begin{aligned} &\min \max_{s'\in S_f} R_{s'} \quad (50)\\ &\quad s.t.\ (2), (3), (6), (29) - (31),\\ &\quad R_{s'} = \left[\sum_{k\in N} f_k^{s'} x_{kk} + \sum_{s\in S_w} p_s \left(\sum_{i\in N}\sum_{k\in N} \chi d_{ik} O_i^s x_{ik} + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y_{kl}^{is} + \sum_{i\in N}\sum_{k\in N} \delta d_{ki} D_i^s x_{ik} \right) \right] - Z_{s'}^* \quad s' \in S_f \quad (51) \end{aligned}$$

As before, the above problem can be easily linearized by defining a variable R and setting it greater or equal to all $R_{s'},\ s' \in S_f$.

In practice, the model above can be a large-scale MILP model and thus the possibility of solving it to optimality using a general solver may strongly depend on the number of scenarios for the uncertainty and, naturally, also on the number of nodes in the network.

5.2. Multiple allocation

For multiple allocation version, we can follow exactly the same reasoning as in the previous section. Accordingly, following the information provided in Sub Sections 3.2 and 4.2, for each scenario $s' \in S_f$, we can consider the following stochastic multiple allocation hub location problem:

$$\begin{aligned} &\min \sum_{k\in N} f_k^{s'} x_k + \sum_{s\in S_w} p_s \left(\sum_{i\in N}\sum_{k\in N} \chi d_{ik} u_{ik}^s + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y_{kl}^{is} + \sum_{i\in N}\sum_{l\in N}\sum_{j\in N} \delta d_{lj} v_{lj}^{is} \right) \quad (52)\\ &\quad s.t.\ (15), (42) - (48). \end{aligned}$$

Denote by $Z_{s'}^*$ the optimal value of the above problem. We can now propose a robust-stochastic model for the problem:

$$\begin{aligned} &\min \max_{s'\in S_f} R_{s'} \quad (53)\\ &\quad s.t.\ (15), (42) - (48),\\ &\quad R_{s'} = \left[\sum_{k\in N} f_k^{s'} x_k + \sum_{s\in S_w} p_s \left(\sum_{i\in N}\sum_{k\in N} \chi d_{ik} u_{ik}^s + \sum_{i\in N}\sum_{k\in N}\sum_{l\in N} \alpha d_{kl} y_{kl}^{is} + \sum_{i\in N}\sum_{l\in N}\sum_{j\in N} \delta d_{lj} v_{lj}^{is} \right) \right] - Z_{s'}^* \quad s' \in S_f \quad (54) \end{aligned}$$

Again, the linearization of this problem is straightforward by defining a new variable R and setting it greater or equal to all $R_{s'},\ s' \in S_f$.

5.3. Computational analysis

In this section, we analyze the results with the single and multiple allocation robust-stochastic hub location models dealing with uncertainty both in the set-up costs and in the demands. Similar to our previous analyses, we use the CAB data set. All the parameter settings comply with the values presented in Sections 3.3 and 4.3. We use the same scenarios as we described in Section 3.3 for the set-up costs, denoted by $S_f^1,\ S_f^2,\ S_f^3,\ S_f^4$, and S_f^5, and the same scenarios described in Section 4.3 for the stochastic demands: $S_w^1,\ S_w^2,\ S_w^3,\ S_w^4$, and S_w^5.

For both the single and multiple allocation versions of the problems, initially, we solve the stochastic models under each set-up cost scenario, and then solve the robust-stochastic models. All the instances are solved to optimality within at most nineteen minutes by using CPLEX version 11.2. The results are presented in Tables 5 and 6.

The optimal locations of the hub nodes in Tables 5 and 6 differ a lot under different fixed cost scenarios. This shows that the stochastic models are also sensitive to the changes in the set-up costs. In general, the effect of uncertainty in the set-up costs on the resulting solutions are higher than the effect of uncertainty in the demand.

Additionally, we want to elaborate the isolated and combined effects of different sources of uncertainty on the resulting hub locations. For this analysis, we summarize the optimal hub locations that we obtained through Tables 1–6 in Table 7.

Observe from Table 7 that optimal hub locations tend to differ under different sources of uncertainty. Even though the optimal solutions are different, there are some hub nodes which are common in the results. For example, node 12 is almost

Table 5
Results with the single allocation robust-stochastic model.

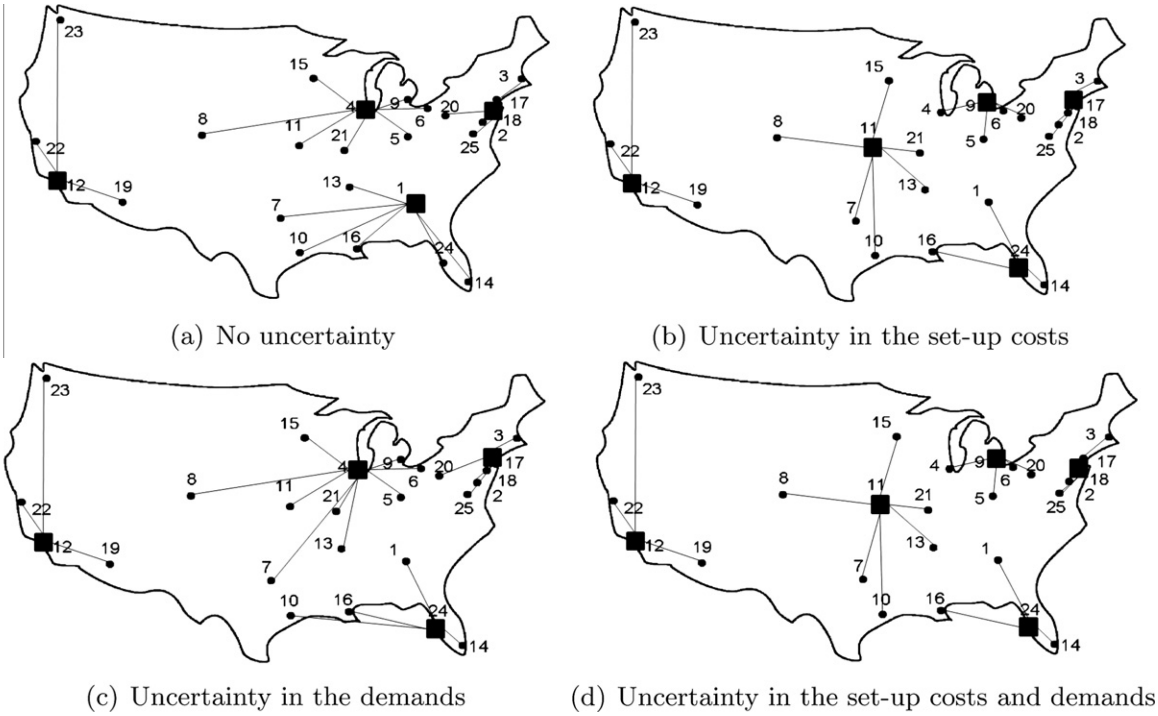
	$\alpha = 0.2$		$\alpha = 0.4$		$\alpha = 0.6$		$\alpha = 0.8$	
	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.
Base case: $f_k = mean,\ w_{ij} = \bar{w}_{ij}$	968.83	4,7,12,17,24	1133.55	1,4,12,18	1280.10	1,4,12,18	1412.10	2,4,12
Stochastic model with S_f^1	896.79	9,11,17,22,24	1057.70	9,11,12,17,24	1213.36	9,11,12,17,24	1361.84	9,11,17,19,24
Stochastic model with S_f^2	881.48	4,12,14,16,17	1046.81	4,12,14,16,17	1204.97	4,12,14,18	1322.04	5,8,14,18
Stochastic model with S_f^3	915.04	6,11,18,22,24	1066.83	6,13,18,22,24	1216.44	6,13,18,22,24	1355.91	6,13,18,22
Stochastic model with S_f^4	891.71	4,12,17,24	1049.79	2,4,12,13,24	1190.57	1,2,4,12	1316.24	2,4,12
Stochastic model with S_f^5	883.24	9,12,13,14,17	1040.07	9,12,13,14,17	1193.09	5,12,14,17	1338.20	5,12,17
Robust-stochastic model	65.78	9,11,12,17,24	66.33	9,11,12,18,24	63.02	5,12,18,24	35.81	2,5,8

Table 6
Results with the multiple allocation robust-stochastic model.

	$\alpha = 0.2$		$\alpha = 0.4$		$\alpha = 0.6$		$\alpha = 0.8$	
	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.	Obj.	Hub loc.
Base case: $f_k = mean,\ w_{ij} = \bar{w}_{ij}$	962.34	4,7,12,17,24	1097.18	4,12,18,24	1203.77	12,18,21	1271.61	12,18,21
Stochastic model with S_f^1	896.11	9,11,17,22,24	1034.43	9,11,12,17,24	1149.64	9,11,12,17,24	1234.79	9,17,19,24
Stochastic model with S_f^2	873.65	4,12,14,16,17	1019.13	4,12,14,16,18	1131.95	4,12,14,18	1206.70	5,8,18
Stochastic model with S_f^3	911.54	6,11,18,22,24	1045.07	6,11,18,22,24	1157.07	6,13,18,22	1214.49	18,21,22
Stochastic model with S_f^4	882.47	4,12,17,24	1011.92	2,4,12,24	1114.51	2,4,12,24	1186.84	2,4,12
Stochastic model with S_f^5	876.57	6,12,14,17,21	1008.18	6,12,14,17,21	1125.99	5,12,14,17	1197.62	5,12,17
Robust-stochastic model	66.46	4,17,22,24	64.30	4,12,18,24	57.68	4,12,18,24	50.43	2,12,21

Table 7
The effects of different sources of uncertainty on the optimal hub locations.

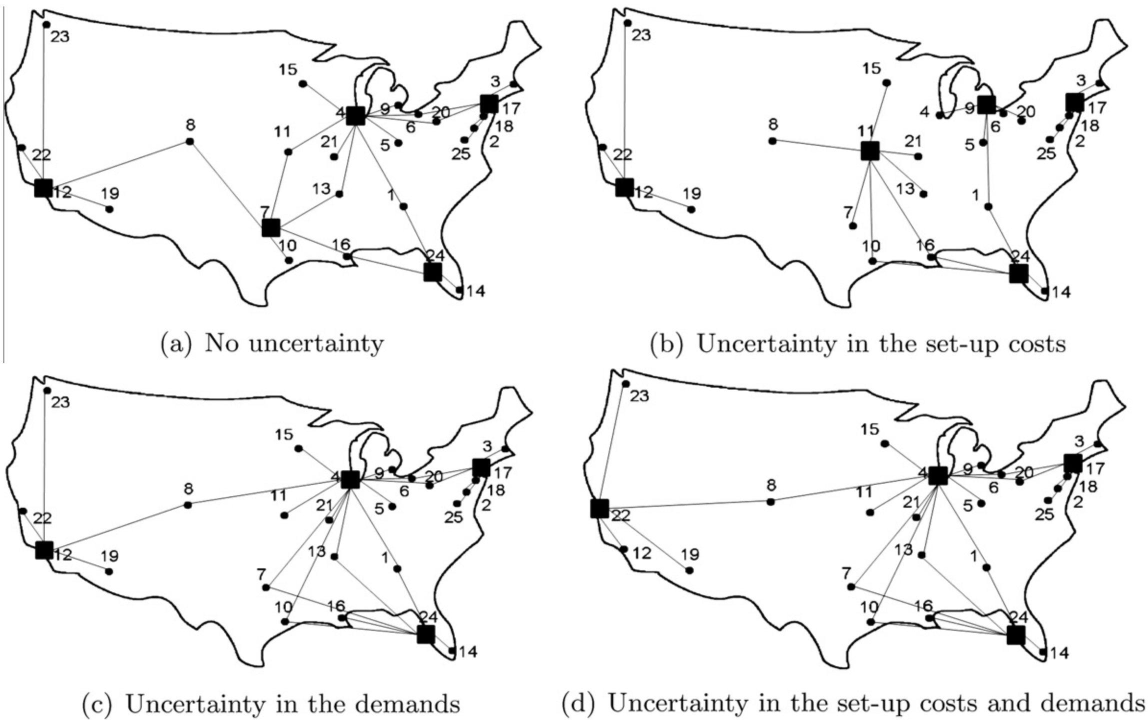
Sources of uncertainty	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
Single allocation				
No uncertainty	4,7,12,17,24	1,4,12,18	1,4,12,18	2,4,12
Set-up costs	9,11,12,17,24	9,11,12,17,24	5,12,18,24	2,5,8
Demands	4,12,17,24	4,12,17,24	4,12,14,18	2,5,12
Set-up costs and demands	9,11,12,17,24	9,11,12,18,24	5,12,18,24	2,5,8
Multiple allocation				
No uncertainty	4,7,12,17,24	4,12,18,24	12,18,21	12,18,21
Set-up costs	9,11,12,17,24	4,12,18,24	4,12,18,24	2,12,21
Demands	4,12,17,24	4,12,17,24	4,12,18,24	12,18,21
Set-up costs and demands	4,17,22,24	4,12,18,24	4,12,18,24	2,12,21



(a) No uncertainty (b) Uncertainty in the set-up costs

(c) Uncertainty in the demands (d) Uncertainty in the set-up costs and demands

Fig. 1. Results with the single allocation problems when $\alpha = 0.4$.



(a) No uncertainty (b) Uncertainty in the set-up costs

(c) Uncertainty in the demands (d) Uncertainty in the set-up costs and demands

Fig. 2. Results with the multiple allocation problems when $\alpha = 0.2$.

always chosen as a hub under different sources of uncertainty. Similarly, either node 17 or node 18 is usually present in the hub set. These results show that some hub locations are robust to the changes in the data. This is mainly due to geographical locations of the nodes in the CAB data set. In addition to the set-up costs and demands, geographical location also has a major impact on determining the optimal locations of the hubs. Thus, we want to visually observe the effects of uncertainty on the resulting solutions as well. We take an instance from the single and the multiple allocation versions of the problems and demonstrate the resulting solutions on the map.

Fig. 1 presents the solutions obtained with different sources of uncertainty with the single allocation problem when $\alpha = 0.4$. In all of the solutions in Fig. 1, there is a hub located at node 12. In addition, there is a hub located at the north-eastern coast, at node 17 or 18. In all the solutions, a hub is located at node 1 or node 24. When there is any uncertainty in the set-up costs, an additional hub is located. Instead of locating a single hub at node 4, with uncertainty in the set-up costs hubs are located at nodes 9 and 11.

Fig. 2 demonstrates the resulting solutions with the multiple allocation problem when $\alpha = 0.2$. In all of the instances, a hub is located at nodes 17 and 24. In all of the instances except in Fig. 2b a hub is located at node 4. Similar to the results presented in Fig. 1, with uncertainty in the set-up costs, two hubs are located at nodes 9 and 11 instead of locating a single hub at node 4. In all of the instances except in Fig. 2d a hub is located at node 12. When there is uncertainty both in the set-up costs and demands, the hub located at node 12 is moved to node 22. When there is no uncertainty or when there is uncertainty only in the set-up costs a total of five hubs are located. However, with the consideration of uncertainty in the demand one less hub is required to provide the optimal service for this specific instance.

Lastly, note that there are some common hub locations in the results presented in Figs. 1 and 2. In addition to different sources of uncertainty, the decision maker can make use of the results presented in Table 7 to identify hub locations which are robust to the economies of scale parameter and also to the allocation decisions.

6. Conclusions

In this paper, we studied hub location problems under uncertainty in the set-up costs and in the demands. We presented generic models capturing these different sources of uncertainty for the single and the multiple allocation versions of the problems. We performed extensive computational analysis with more than 150 instances on the CAB data set. We analyzed the changes in the optimal hub locations driven by the different sources of uncertainty.

Our results show that the structure of the solutions changes when uncertainty is considered. We were able to show that the hub locations resulting from models ignoring uncertainty in the data are not optimal and that the total costs may increase a lot if the decision maker does not anticipate uncertainty. In the computational tests performed, the impact of considering uncertainty in the set-up costs seems to he higher than the impact of considering uncertainty in the demands. Note, however, that this conclusion cannot be generalized as the impact of the uncertainty in the solutions is strongly instance dependent. Finally, our analysis also shows that simultaneously considering uncertainty in the set-up costs and in the demands led in many cases to solutions that are different from the ones obtained when no uncertainty is considered or when only one source of uncertainty is considered. This shows that the optimal solution is clearly sensitive to the inclusion of uncertainty in the model and, more specifically to the components of the data that exhibit uncertainty. Accordingly, a modeling framework embedding uncertainty is clearly worth considering in the decision making process.

In this paper, we considered a simple setting to demonstrate the proposed methodology. Nevertheless, the methods presented in this paper can be extended to more complex hub location problems such as capacitated problems and problems with hub network design decisions.

Acknowledgements

The authors appreciate the constructive criticism of the anonymous referee, whose feedback helped us in improving the paper. The first author is supported from the Scientific and Technological Research Council of Turkey (TÜBİTAK); third author is supported by the *Portuguese Science Foundation*, POCTI-ISFL-1-152.

References

Abdinnour-Helm, S., 1998. A hybrid heuristic for the uncapacitated hub location problem. European Journal of Operational Research 106 (2), 489–499.
Alumur, S., Kara, B.Y., 2008. Network hub location problems: the state of the art. European Journal of Operational Research 190 (1), 1–21.
Alumur, S., Kara, B.Y., 2009. A hub covering network design problem for cargo applications in Turkey. Journal of the Operational Research Society 60 (10), 1349–1359.
Alumur, S.A., Kara, B.Y., Karasan, O.E., 2009. The design of single allocation incomplete hub networks. Transportation Research Part B 43 (10), 936–951.
Aykin, T., 1994. Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. European Journal of Operational Research 79 (3), 501–523.
Birge, J.R., Louveaux, F., 1997. Introduction to Stochastic Programming. Springer.
Boland, N., Krishnamoorthy, M., Ernst, A.T., Ebery, J., 2004. Preprocessing and cutting for multiple allocation hub location problems. European Journal of Operational Research 155 (3), 638–653.
Calik, H., Alumur, S.A., Kara, B.Y., Karasan, O.E., 2009. A tabu-based heuristic for the hub covering problem over incomplete hub networks. Computers & Operations Research 36 (12), 3088–3096.
Campbell, J.F., 1994. Integer programming formulations of discrete hub location problems. European Journal of Operational Research 72 (2), 387–405.
Campbell, J.F., 1996. Hub location and the *p*-hub median problem. Operations Research 44 (6), 1–13.
Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005a. Hub arc location problems: Part I – introduction and results. Management Science 51 (10), 1540–1555.
Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2005b. Hub arc location problems: Part II – formulations and optimal algorithms. Management Science 51 (10), 1556–1571.

Contreras, I., Cordeau, J.-F., Laporte, G., in press. Exact solution of large-scale hub location problems with multiple capacity levels. Transportation Science.
Contreras, I., Fernández, E., Marín, A., 2009. Tight bounds from a path based formulation for the tree of hub location problem. Computers & Operations Research 36 (12), 3117–3127.
Contreras, I., Fernández, E., Marín, A., 2010. The tree of hubs location problem. European Journal of Operational Research 202 (2), 390–400.
Contreras, I., Cordeau, J.-F., Laporte, G., 2011. Stochastic uncapacitated hub location. European Journal of Operational Research 212 (3), 518–528.
Contreras, I., Cordeau, J.-F., Laporte, G., 2012. Benders decomposition for large-scale uncapacitated hub location. Operations Research. doi:10.1287/opre.1110.0965.
Correia, I., Nickel, S., da Gama, F.S., 2010. The capacitated single-allocation hub location problem revisited: a note on a classical formulation. European Journal of Operational Research 207 (1), 92–96.
Costa, M., Captivo, M., Clímaco, J., 2008. Capacitated single allocation hub location problem – a bi-criteria approach. Computers & Operations Research 35 (11), 3671–3695.
Ebery, J., Krishnamoorthy, M., Ernst, A., Boland, N., 2000. The capacitated multiple allocation hub location problem: formulations and algorithms. European Journal of Operational Research 120 (3), 614–631.
Ernst, A.T., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation *p*-hub median problem. Location Science 4 (3), 139–154.
Ernst, A.T., Krishnamoorthy, M., 1998a. Exact and heuristic algorithms for the uncapacitated multiple allocation *p*-hub median problem. European Journal of Operational Research 104 (1), 100–112.
Ernst, A.T., Krishnamoorthy, M., 1998b. An exact solution approach based on shortest-paths for *p*-hub median problems. Informs Journal on Computing 10 (2), 149–162.
Ernst, A.T., Krishnamoorthy, M., 1999. Solution algorithms for the capacitated single allocation hub location problem. Annals of Operations Research 86 (0), 141–159.
Hamacher, H., Labbé, M., Nickel, S., Sonneborn, T., 2004. Adapting polyhedral properties from facility to hub location problems. Discrete Applied Mathematics 145 (1), 104–116.
Kleywegt, A.J., Shapiro, A., de Mello, T.H., 2001. The sample average approximation method for stochastic discrete optimization. SIAM Journal of Optimization 12 (2), 479–502.
Kouvelis, P., Yu, G., 1997. Robust Discrete Optimization and Its Applications. Kluwer Academic Publishers.
Labbé, M., Yaman, H., 2004. Projecting flow variables for hub location problems. Networks 44 (2), 84–93.
Labbé, M., Yaman, H., Gourdin, E., 2005. A branch and cut algorithm for the hub location problems with single assignment. Mathematical Programming 102 (2), 371–405.
Lium, A.-G., Crainic, T.G., Wallace, S.W., 2009. A study of demand stochasticity in service network design. Transportation Science 43 (2), 144–157.
Marianov, V., Serra, D., 2003. Location models for airline hubs behaving as M/D/c queues. Computers & Operations Research 30 (7), 983–1003.
Marin, A., 2005. Formulating and solving splittable capacitated multiple allocation hub location problems. Computers & Operations Research 32 (12), 3093–3109.
Mayer, G., Wagner, B., 2002. Hublocator: an exact solution method for the multiple allocation hub location problem. Computers & Operations Research 29 (6), 715–739.
Nickel, S., Schobel, A., Sonneborn, T., 2001. Hub location problems in urban traffic networks. In: Niittymaki, J., Pursula, M. (Eds.), Mathematics Methods and Optimization in Transportation Systems. Kluwer Academic Publishers, pp. 1–12.
O'Kelly, M., 1986. Activity levels at hub facilities in interacting networks. Geographical Analysis 18 (4), 343–356.
O'Kelly, M., 1987. A quadratic integer problem for the location of interacting hub facilities. European Journal of Operational Research 32 (3), 393–404.
O'Kelly, M., Bryan, D., Skorin-Kapov, D., Skorin-Kapov, J., 1996. Hub network design with single and multiple allocation: a computational study. Location Science 4 (3), 125–138.
OR-Library, 2011. <http://people.brunel.ac.uk/∼mastjjb/jeb/info.html>.
Pirkul, H., Schilling, D., 1998. An efficient procedure for designing single allocation hub and spoke systems. Management Science 44 (12), 235–242.
Sasaki, M., Fukushima, M., 2003. On the hub-and-spoke model with arc capacity constraints. Journal of the Operations Research Society of Japan 46 (4), 409–428.
Sim, T., Lowe, T.J., Thomas, B.W., 2009. The stochastic *p*-hub center problem with service-level constraints. Computers & Operations Research 36 (12), 3166–3177.
Skorin-Kapov, D., Skorin-Kapov, J., 1994. On tabu search for the location of interacting hub facilities. European Journal of Operational Research 73 (3), 502–509.
Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, M., 1996. Tight linear programming relaxations of uncapacitated *p*-hub median problems. European Journal of Operational Research 94 (3), 582–593.
Snyder, L., 2006. Facility location under uncertainty: a review. IIE Transactions 38 (7), 537–554.
Yaman, H., 2008. Star *p*-hub median problem with modular arc capacities. Computers & Operations Research 35 (9), 3009–3019.
Yaman, H., 2011. Allocation strategies in hub networks. European Journal of Operational Research 211 (3), 442–451.
Yang, T.-H., 2009. Stochastic air freight hub location and flight routes planning. Applied Mathematical Modelling 33 (12), 4424–4430.
Yoon, M., Current, J., 2008. The hub location and network design problem with fixed and variable arc costs. Journal of the Operational Research Society 59 (1), 80–89.