A Modeling Framework for Stochastic Multi-Period Capacitated Multiple Allocation Hub Location

Isabel Correia [1], Stefan Nickel [2,3], Francisco Saldanha-da-Gama [4]

[1] Departamento de Matemática / CMA, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisbon, Portugal

[2] Institute of Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[3] Department of Logistics and Supply Chain Optimization, Research Center for Information Technology (FZI), Karlsruhe, Germany

[4] DEIO / CMAF-CIO, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

{isc@fct.unl.pt, stefan.nickel@kit.edu, fsgama@ciencias.ulisboa.pt}

Abstract. We propose a two-stage stochastic programming modeling framework for multi-period multiple allocation hub location under uncertainty. A discretized planning horizon is considered and stochasticity is assumed for the flows to be routed through the network. When uncertainty can be described by a discrete random vector with a finite support it is possible to derive the extensive form of the deterministic equivalent. However, this results in a large-scale mixed-integer linear programming model that nonetheless can be enhanced using several families of valid inequalities. Computational tests performed using benchmark data are reported and show that the new sets of valid inequalities are able to provide a good polyhedral description of the feasibility set, which is of relevance

Keywords: Hub Location, Multiple Allocation, Multi-Period, Two-Stage Stochastic Programming, Valid Inequalities.

1 Introduction

In a hub location problem a set of nodes is given such that a flow must be sent between each pair of nodes. By imposing some structure to the underlying network it may be possible to route the flows between origin-destination (O-D) nodes in an more “efficient” way. A common structure is the one induced by the selection of some nodes to become hubs, i.e., to receive the flow (e.g., mail, passengers) and redistribute it (to their destination or to other hubs). The links between hubs are usually transversed by large amounts of flow, which makes it possible to take advantage from economies of scale in terms of transportation costs.
In this work we investigate a stochastic multi-period multiple allocation hub location problem. We consider a discretized and finite planning horizon. The decisions to be made concern (i) when and where to install new hubs and their initial capacity, (ii) when and where to expand capacity for existing hubs, and (iii) how to route the flow between O-D pairs in each period. We make several assumptions:

- The potential locations for hubs are a subset of the nodes underlying the problem.
- At the beginning of each time period it is possible to open new hubs. When this happens it is necessary to decide how many modules the new hub will get. In the beginning of subsequent periods it is possible to install additional modules although a limit is assumed for the maximum number of modules allowed in each location.
- A hub must process its own outbound flow and distribute its own inbound flow.
- Capacities are modular, i.e., the capacity of a hub at some point in time is fully determined by the number of modules installed in the hub so far.
- The capacity of each hub refers to the maximum flow it can process.
- During the planning horizon hubs can neither be closed nor their capacities reduced, i.e., we are considering a pure phase-in problem.

- Each non-hub node can be allocated to more than one hub in each time period. Such allocations can change during the planning horizon.
- Uncertainty is associated with the flows in the different periods of the planning horizon.
- Uncertainty can be captured by a finite set of scenarios, each having some occurrence probability.
- A risk neutral attitude is assumed for the decision maker. This means that the current value of future assets will be captured by expected values.

The goal of the problem is to minimize the total cost, which includes (i) the setup of cost for installing new hubs and their initial operating capacity, (ii) the cost for expanding the capacity of existing hubs, (iii) the cost for operating hubs, and (iv) the cost for routing the flow through the network.
To the best of the authors' knowledge, no work has been published within the area of hub location that casts stochasticity in a multi-period setting. Nevertheless, some works can be found dealing with these aspects separately: [4], [7], and [9] study multi-period hub location problems; stochastic hub location problems have been investigated in [3], [8], [15], [16], [17], and [18].
The literature addressing hub location problems with modular capacities is also scarce. Modular capacities in the context of hub location have been previously considered in [4], [10], and [12].
We formulate our problem using a two-stage stochastic programming modeling framework: the first stage captures the strategic decisions and consists of defining a plan for locating the hubs and setting up their initial capacity for the entire planning horizon; the second stage captures the tactical and operational decisions, that is, capacity expansion of existing hubs and flow routing. We note that in both stages of the modeling framework we are defining a multi-period plan: in the first stage we define a multi-period plan for locating hubs and their capacities; in the second stage we will be designing a a multi-period plan for expanding capacities and for serving demand. The reasoning underlying this two-stage framework has to do with the fact that strategic decisions often require time to be implemented whereas tactical and operational decisions can be more easily adjusted (namely when more information about the flows becomes available). We also point out that the application of a two-stage stochastic modeling framework to stochastic multi-period problems has been considered by some authors for other problems (see, e.g., [1] and [2]).

2 Optimization Model

As it was already mentioned in the previous section, we propose a two-stage modeling framework for our problem. Accordingly, before uncertainty is revealed, there is a *here-and-now* decision to make. Afterwards, uncertainty is disclosed and a recourse action is taken.
In our particular case, the first stage decision concerns the location of the hubs over the planning horizon as well as their initial capacity. In the second decision level we include decisions about the allocation of the non-hubs to the hubs, the value of the flow variables in all time periods of the planning horizon, and the capacity expansion for existing hubs.
In order to formulate an optimization model, we take as a starting point the well-known formulation proposed in [6] for static and deterministic multiple allocation hub location problems.
Next we introduce some notation that will be used throughout the paper.

Deterministic information:

N, set of nodes;
R, set of potential locations for the hubs ($R \subseteq N$);
T, set of time periods in the planning horizon;

Γ, capacity of a module;
Q_k, maximum number of modules that can be installed in a hub at location $k \in R$;

d_{ij}, distance between nodes $i \in N$ and $j \in N$;
χ^t, collection cost in period $t \in T$, i.e., cost per unit of distance of one unit of flow that goes from a non-hub node to a hub in period t;
δ^t, distribution cost in period t, i.e., cost per unit of distance of one unit of flow that goes from a hub to a non-hub node in period $t \in T$;

α^t, transfer cost in period t, i.e., cost per unit of flow and per unit of distance between hubs in period $t \in T$;

g_k^{qt}, total cost resulting from setting up the initial capacity of a hub to be located at $k \in R$ in period $t \in T$ with q modules. This cost now includes the operating cost for that capacity until the end of the planning horizon;

h_k^{qt}, cost to be paid from period $t \in T$ until the end of the planning horizon when the capacity of (an existing) hub $k \in R$ is expanded using q additional modules.

Stochastic information

S, Finite set of scenarios describing uncertainty;

p_s, probability that scenario $s \in S$ occurs, $\sum_{s\in S} p_s = 1$;

w_{ijs}^t, flow to be sent from node $i \in N$ to node $j \in N$ in period $t \in T$, and scenario $s \in S$;

O_{is}^t, total flow originated at node $i \in N$ in period $t \in T$ and scenario $s \in S$, $O_{is}^t = \sum_{j\in N} w_{ijs}^t$;

D_{is}^t, total flow destined to node $i \in N$ in period $t \in T$ and scenario $s \in S$, $D_{is}^t = \sum_{j\in N} w_{jis}^t$.

First stage decision variables:

$$u_k^t = \begin{cases} 1, & \text{if a hub is installed at } k \in R \text{ in period } t \in T, \\ 0, & \text{otherwise.} \end{cases} \qquad (k \in R, t \in T)$$

$$z_k^{qt} = \begin{cases} 1, & \text{if hub } k \text{ receives } q \text{ modules in period } t, \\ 0, & \text{otherwise.} \end{cases} \qquad (k \in R, q = 1, ..., Q_k, t \in T)$$

Recourse decision variables:

$$r_{ks}^{qt} = \begin{cases} 1, & \text{if hub } k \text{ receives } q \text{ additional modules in period } t \text{ and scenario } s, \\ 0, & \text{otherwise.} \end{cases} \qquad (k \in R, q = 1, ..., Q_k - 1, t \in T \setminus \{1\}, s \in S)$$

x_{iks}^t = amount of flow with origin at $i \in N$ that is collected at hub $k \in R$ in period $t \in T$ and scenario $s \in S$.

y_{kls}^{it} = amount of flow with origin at $i \in N$ that is collected at hub $k \in R$ and is distributed by hub $l \in R$ in period $t \in T$ and scenario $s \in S$.

v_{ljs}^{it} = amount of flow with origin at $i \in N$ destined to node $j \in N$ that is distributed by hub $l \in N$ in period $t \in T$ and scenario $s \in S$.

Using all the information above introduced, we can formulate the extensive for of the deterministic equivalent of our problem—hereafter denoted by DE—as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{k\in R}\sum_{t\in T}\left(f_k^t u_k^t + \sum_{q=1}^{Q_k} g_k^{qt} z_k^{qt}\right) \\ & + \sum_{s\in S} p_s \Bigg[\sum_{i\in N}\sum_{k\in R}\sum_{t\in T}\chi^t d_{ik} x_{iks}^t + \sum_{i\in N}\sum_{k\in R}\sum_{l\in R}\sum_{t\in T}\alpha^t d_{kl} y_{kls}^{it} \\ & \qquad + \sum_{i\in N}\sum_{l\in R}\sum_{j\in N}\sum_{t\in T}\delta^t d_{lj} v_{ljs}^{it} + \sum_{k\in R}\sum_{q=1}^{Q_k-1}\sum_{t\in T\setminus\{1\}} h_k^{qt} r_{ks}^{qt}\Bigg], \end{aligned} \tag{1}$$

$$\text{subject to} \quad \sum_{q=1}^{Q_k} z_k^{qt} = u_k^t, \quad k \in R, t \in T, \tag{2}$$

$$\sum_{t\in T} u_k^t \le 1, \quad k \in R, \tag{3}$$

$$u_k^t \in \{0,1\}, \quad k \in R, t \in T, \tag{4}$$

$$z_k^{qt} \in \{0,1\}, \quad k \in R, q = 1, \ldots, Q_k, t \in T, \tag{5}$$

$$\sum_{k\in R} x_{iks}^t = O_{is}^t, \quad i \in N, t \in T, s \in S, \tag{6}$$

$$\sum_{l \in R} v_{ljs}^{it} = w_{ijs}^{t}, \quad i,j \in N, t \in T, s \in S, \tag{7}$$

$$\sum_{q=1}^{Q_k-1} r_{ks}^{qt} \leq \sum_{\tau=1}^{t-1} u_k^{\tau}, \quad k \in R, t \in T \setminus \{1\}, s \in S, \tag{8}$$

$$\sum_{i \in N} x_{iks}^{t} \leq \Gamma \sum_{q=1}^{Q_k} \sum_{\tau=1}^{t} q z_k^{q\tau} + \Gamma \sum_{q=1}^{Q_k-1} \sum_{\tau=2}^{t} q r_{ks}^{q\tau}, \quad k \in R, t \in T, s \in S, \tag{9}$$

$$\sum_{q=1}^{Q_k} \sum_{t \in T} q z_k^{qt} + \sum_{q=1}^{Q_k-1} \sum_{t \in T \setminus \{1\}} q r_{ks}^{qt} \leq Q_k, \quad k \in R, s \in S, \tag{10}$$

$$\sum_{l \in R, l \neq k} y_{kls}^{it} - \sum_{l \in R, l \neq k} y_{lks}^{it} = x_{iks}^{t} - \sum_{j \in N} v_{kjs}^{it}, \quad i \in N, k \in R, t \in T, s \in S, \tag{11}$$

$$x_{iks}^{t} \leq O_{is}^{t}(1 - \sum_{\tau=1}^{t} u_i^{\tau}), \quad i,k \in R : i \neq k, t \in T, s \in S, \tag{12}$$

$$\sum_{i \in N} v_{ljs}^{it} \leq D_{js}^{t} \sum_{\tau=1}^{t} u_l^{\tau}, \quad j \in N, l \in R, t \in T, s \in S, \tag{13}$$

$$v_{ljs}^{it} \leq D_{js}^{t}(1 - \sum_{\tau=1}^{t} u_j^{\tau}), \quad i \in N, l,j \in R : l \neq j, t \in T, s \in S, \tag{14}$$

$$r_{ks}^{qt} \in \{0,1\}, \quad k \in R, q = 1, \ldots, Q_k - 1, t \in T \setminus \{1\}, s \in S, \tag{15}$$

$$x_{iks}^{t} \geq 0, \quad i \in N, k \in R, t \in T, s \in S, \tag{16}$$

$$y_{kls}^{it} \geq 0, \quad i \in N, k,l \in R, t \in T, s \in S, \tag{17}$$

$$v_{ljs}^{it} \geq 0, \quad i,j \in N, l \in R, t \in T, s \in S. \tag{18}$$

In the above model, the objective function (1) evaluates the total expected cost which consists of the first stage cost (installing the hubs and their initial capacities) plus the expected cost associated with the recourse decisions (capacity expansion for existing hubs and flow routing). Constraints (2) ensure that in each time period when a hub is installed there must must be a certain number of modules defining its initial capacity. Constraints (3) state that one hub can be installed at most once during the planning horizon. Constraints (4) and (5) are the domain constraints for the first stage decision variables.

The constraints associated with the second stage problem (one group for each scenario $s \in S$) start with (6) ensuring that in each time period the flow originated in each node is routed via at least one hub; in turn, (7) ensure that in each time period the flow between each pair of nodes is distributed via at least one hub. Inequalities (8) state that the expansion of the capacity of a hub in a time period is possible only if the hub had been previously installed. Relations (9) are the capacity constraints for the hubs. They ensure that for each hub and in each time period the incoming flow from non-hubs as well as the flow originated at the hub cannot exceed the operating capacity. In each potential location for the hubs there is a maximum number of modules that can be installed. This is stated in constraints (10). Relations (11) are the flow conservation and divergence constraints. Inequalities (12) impose that a hub must collect its own outbound flow. By constraints (13) we ensure that the flow destined to a node can be distributed only by an open hub whereas (14) guarantees that a hub distributes its own inbound flow. Finally, (15)–(18) are domain constraints.

DE is a large-scale mixed-integer linear programming model even for tiny instances of the problem. This fact motivates the developments that we present next.

3 Valid Inequalities

In order to enhance the extensive form of the deterministic equivalent presented in the previous section, we propose several sets of valid inequalities. Below we present a brief synthesis of the deep analysis performed in [13].

The first set of valid inequalities is generated by the Chávatal-Gomory rounding method [1] and is derived from constraints (10) that establish an upper bound on the total number of modules installed in each potential location and for each scenario. The reasoning we use next has been also considered among other authors (see [14] for concentrator location problems and [11] for the variable size bin packing problem).

Result: The following inequalities are valid for DE:

$$\sum_{t\in T}\sum_{q=1}^{Q_k}\left\lfloor\frac{q}{p}\right\rfloor z_k^{qt} + \sum_{t\in T\setminus\{1\}}\sum_{q=1}^{Q_k-1}\left\lfloor\frac{q}{p}\right\rfloor r_{ks}^{qt} \le \left\lfloor\frac{Q_k}{p}\right\rfloor \sum_{t\in T} u_k^t, \qquad k\in R,\ s\in S,\ p=1,\ldots,Q_k. \tag{19}$$

◇

A second set of valid inequalities is derived from the assumption that each hub processes its outbound flow combined with the rounding approach used in the previous result. The corresponding result can be proved.

Result: The following inequalities are valid for DE:

$$\sum_{q=1}^{Q_k}\left\lceil\frac{q}{p}\right\rceil z_k^{qt} + \sum_{\tau=t+1}^{|T|}\sum_{q=1}^{Q_k-1}\left\lceil\frac{q}{p}\right\rceil r_{ks}^{q\tau} \ge \left\lceil\frac{\Theta_{ks}^t}{p}\right\rceil u_k^t, \qquad k\in R,\ t\in T,\ s\in S,\ p=1,\ldots,\Theta_{ks}^t, \tag{20}$$

where Θ_{ks}^t denotes the minimum number of modules that need to be installed in location $k\in R$ from period $t\in T$ until the end of the planning horizon in some scenario $s\in S$

◇

In order to derive the third set of valid inequalities, we can compute a lower bound, Δ_s^t, on the total number of modules that must be operating in period $t\in T$ under scenario $s\in S$. From this quantity we can write a set of constraints ensuring that in period t and under scenario s at least Δ_s^t modules will be operating. From these inequalities and using again a similar reasoning as presented above we can prove the following result:

Result: The following inequalities are valid for DE:

$$\sum_{k\in R}\sum_{q=1}^{Q_k}\sum_{\tau=1}^{t}\left\lceil\frac{q}{p}\right\rceil z_k^{qt} + \sum_{k\in R}\sum_{q=1}^{Q_k-1}\sum_{\tau=2}^{t}\left\lceil\frac{q}{p}\right\rceil r_{ks}^{q\tau} \ge \left\lceil\frac{\Delta_s^t}{p}\right\rceil, \qquad t\in T,\ s\in S,\ p=1,\ldots,\Delta_s^t, \tag{21}$$

◇

Finally, we propose a fourth set of valid inequalities (of a different type) as follows:

$$x_{iks}^t \le O_{is}^t \sum_{\tau=1}^{t} u_k^t, \qquad i\in N,\ k\in R,\ s\in S,\ t\in T. \tag{22}$$

These inequalities ensure that if no hub is installed in location $k\in R$ until period $t\in T$, then the associated x-variables should be set to zero. We note that these constraints are implied by constraints (2), (8), and (9) together. However, despite being redundant for model DE, adding them to the formulation may render an additional improvement to the linear relaxation bound.

4 Computational Results

We tested the models discussed in the previous sections using a set of instances generated from the well-known CAB data set (available in [5]). We considered instances with $|N|\in\{15,20,25\}$, $|R|\in\{10,15,20,25\}$ ($|R|\le|N|$), $|T|=3$, $|S|=5$, and $Q_k=5$, $k\in R$. Furthermore, we assume that each scenario represents a trend for the flows. Starting from the CAB data (with flows scaled to add 1) we considered 5 possibilities

[1] Suppose we have a valid inequality, $\sum_{j=1}^n a_j x_j \le b$, for a polyhedron P given by the linear relaxation of an integer program (we are assuming that variables x_j can only take non-negative integer values). Then, the inequality $\sum_{j=1}^n \lfloor a_j\rfloor x_j \le b$ is also valid for P. The left hand side is integer for all feasible solution $(x_1,\ldots,x_n)\in X$ where X is the feasibility set of the integer program. Thus, $\sum_{j=1}^n \lfloor a_j\rfloor x_j \le \lfloor b\rfloor$ is a valid inequality for X.

for this trend: a progressive increase in the flows of 6%,12%,18%,24%,and 30%. Finally, as it is usually done with the CAB data set, we considered $\alpha^t \in \{0.2, 0.4, 0.6, 0.8\}$, $t \in T$. Therefore, for each combination of N and R we get 4 instances.
Below we present results for two models: model *DE* and model *DE* enhanced with the 4 sets of valid inequalities proposed in the previous section. The models were implemented using IBM ILOG Concert Technology with IBM Cplex 12.6. The experiments were conducted on a 64-bit machine running ubuntu Operating System, with a processor INTEL XEON E5-2650V3 10CORE 2.3GHZ with 32GB RAM. A CPU time limit of 20 hours was set for each instance.
In Figures 1 and 2 we present results for the instances tested assuming that each module has capacity 0.1. Figure 1 depicts the number of instances solved to optimality within the time limit for each combination of $|N|$ and $|R|$ (we recall that for each such combination 4 instances were generated). By observing this figure we conclude that using the enhanced model can clearly solve more instances to optimality than if we use the base model. Observing Figure 2 we conclude that the valid inequalities proposed are in fact an important

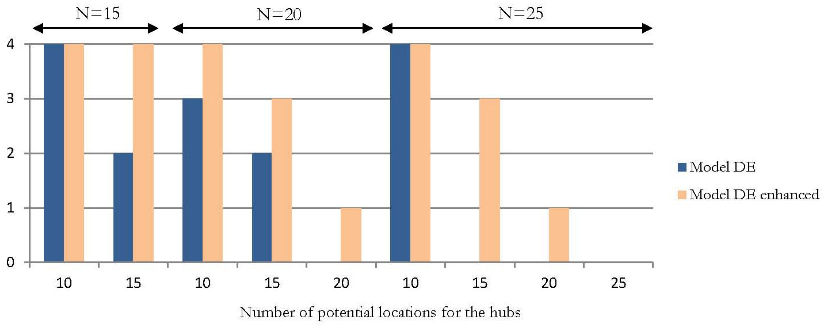

Figure 1: Effectiveness Number of instances solved.

contribution or a better polyhedral description of the feasibility set of the problem we are investigating. Finally, we present a table reporting CPU time (in seconds). Table 1 shows again the superiority of the

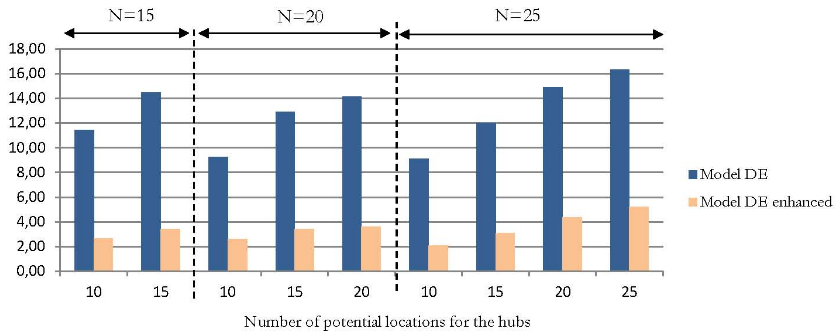

Figure 2: Average linear relaxation gaps (%).

enhanced model over the model without valid inequalities. The CPU time presented is the average CPU time for the instances successfully solved to optimality. In [13] we report extensive results that include the analysis of the relevance of using a stochastic modeling framework for our problem instead of a (simpler) deterministic one.

5 Conclusion

In this work we have investigated a stochastic multi-period capacitated multiple allocation hub location problem. We proposed a two-stage stochastic modeling framework. Assuming that uncertainty can be captured by a finite set of scenarios we were able to derive the extensive form of the deterministic equivalent

Table 1: Average CPU time (seconds) when solving the problem to optimality.

		Model DE		Model DE enhanced	
$\|N\|$	$\|R\|$	# solved	CPU (sec.)	# solved	CPU (sec.)
15	10	4	15226	4	528
	15	2	34275	4	15405
20	10	3	15875	4	4966
	15	2	61757	3	29064
	20	0	—	1	20237
25	10	4	23116	4	2161
	15	0	—	3	30546
	20	0	—	1	37679
	25	0	—	0	—

problem. For the latter model, we proposed several sets of valid inequalities. The computational tests performed show that by using an off-the-shelf solver, the enhanced models make it possibly to solve until optimality instances that could not be solved using the base model. This work represents a first step toward the development of effective models and solution techniques for multi-period stochastic hub location problems.

References

[1] Aghezzaf, E.: Capacity Planning and warehouse location in supply chains with uncertain demands. J Oper Res Soc. 56, 453–462 (2005)

[2] Ahmed, S., Garcia, R.: Dynamic Capacity Acquisition and Assignment under Uncertainty. Ann Oper Res. 124, 267–283 (2004)

[3] Alumur, S., Nickel, S., Saldanha-da-Gama: Hub location under uncertainty. Transport Res B-Meth. 46, 529–543 (2012)

[4] Alumur, S., Nickel, S., Saldanha-da-Gama, F., SeÃ§erdin, Y.: Multi-period hub network design problems with modular capacities. Ann Oper Res. doi: 10.1007/s10479-015-1805-9 (2015)

[5] OR Library: Hub location, http://people.brunel.ac.uk/∼mastjjb/ jeb/orlib/phubinfo.html

[6] Boland, N., Krishnamoorthy, M., Ernst, A.T., Ebery, J.: Preprocessing and cutting for multiple allocation hub location problems. Eur J Oper Res. 155, 638–653 (2004)

[7] Campbell, J.: Locating transportation terminals to serve an expanding demand. Transport Res B-Meth. 24, 173–192 (1990)

[8] Contreras, I., Cordeau, J.-F., Laporte, G.: Stochastic uncapacitated hub location. Eur J Oper Res. 212, 518–528 (2011a)

[9] Contreras, I., Cordeau, J.-F., Laporte, G.: The dynamic uncapacitated hub location problem. Transport Sci. 45, 18–32 (2011b)

[10] Contreras, I., Cordeau, J.-F., Laporte, G.: Exact solution of large-scale hub location problems. Transport Sci. 46, 439–459 (2012)

[11] Correia, I., Gouveia, L., Saldanha-da-Gama, F.: Solving the variable size bin packing problem with discretized formulations. Comput Oper Res. 35, 2103–2113 (2008)

[12] Correia, I., Nickel, S., Saldanha-da-Gama, F.: Single-assignment hub location problem with capacity levels. Transport Res B-Meth. 44, 1047–1066 (2010)

[13] Correia, I., Nickel, S., Saldanha-da-Gama, F.: A stochastic multi-period capacitated multiple allocation hub location problem: formulation and inequalities. Technical Report (submitted to publication).

[14] Gouveia, L., Saldanha-da-Gama, F.: On the capacitated concentrator location problem: a reformulation by discretization. Comput Oper Res. 33, 1242–1258 (2006)

[15] Marianov, V., Serra, D.: Location models for airline hubs behaving as M/D/c queues. Comput Oper Res. 30, 983–1003 (2003)

[16] Qin, Z., Gao, Y.:Uncapacitated p-hub location problem with fixed costs and uncertain flows. J Intell Manuf. doi: 10.1007/s10845-014-0990-8 (2014)

[17] Sim, T., Lowe, T.J., Thomas, B.W.: The stochastic p-hub center problem with service-level constraints. Comput Oper Res. 36, 3166–3177 (2009)

[18] Yang, T.-H.: Stochastic air freight hub location and flight routes planning. Appl Math Model. 33, 4424–4430 (2009)