Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

Stochastic Uncapacitated Hub Location

Ivan Contreras
Jean-François Cordeau
Gilbert Laporte

September 2010

CIRRELT-2010-43

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Stochastic Uncapacitated Hub Location

Ivan Contreras[1,2,3,*], Jean-François Cordeau[1,2], Gilbert Laporte[1,3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)
[2] Canada Research Chair in Logistics and Transportation, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7
[3] Canada Research Chair in Distribution Management, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

Abstract. We study stochastic uncapacitated hub location problems in which uncertainty is associated to demands and transportation costs. We show that the stochastic problems with uncertain demands or dependent transportation costs are equivalent to their associated deterministic expected value problem (EVP), in which random variables are replaced by their expectations. In the case of uncertain independent transportation costs, the corresponding stochastic problem is not equivalent to its EVP and specific solution methods need to be developed. We describe a Monte-Carlo simulation-based algorithm that integrates a sample average approximation scheme with a Benders decomposition algorithm to solve problems having stochastic independent transportation costs. Numerical results on a set of instances with up to 50 nodes are reported.

Keywords. Hub location, stochastic programming, Monte-Carlo sampling, Benders decomposition.

Acknowledgements. This work was partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under grants 227837-09 and 39682-10. This support is gratefully acknowledged.

---

* Corresponding author: Ivan.Contreras@cirrelt.ca

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 2010

1 Introduction

Hub location problems (HLPs) arise in transportation, telecommunication and computer networks, where hub-and-spoke architectures are frequently used to efficiently route commodities between many origin and destination (O/D) pairs. The performance of these networks relies on the use of consolidation, switching, or transshipment points, called *hub facilities*, where flows from several origins are consolidated and rerouted to their destinations, sometimes via another hub. In HLPs the locations of the hubs as well as the paths for sending the commodities have to be determined. Broadly speaking, HLPs consist in locating hubs on a network so as to minimize the total flow cost.

Due to their multiple applications, these problems are receiving increased attention. Solution methods have been developed for several variants of HLPs analogous to well-known discrete facility location problems, such as uncapacitated hub location, p-hub location, p-hub center, and hub covering. For each of these classes of problems, there exist several variants arising from various assumptions, such as hub capacities or a specific topological structure for the hub-and-spoke network. There are two basic assumptions underlying most HLPs. The first is that commodities have to be routed via a set of hubs, and thus paths between O/D pairs include at least one hub facility. The second assumption is that hubs are fully interconnected with more effective, higher volume pathways that enable a discount factor $\tau(0 < \tau < 1)$ to be applied to all transportation costs associated to the commodities routed between a pair of hubs. The reader is refered to Alumur and Kara (2008) and to Campbell et al. (2002) for recent surveys on HLPs.

The location of hub facilities corresponds to long-term strategic decisions which are typically made within an uncertain environment. That is, costs, demands, distances, and other parameters may change after location decisions have been made. Nevertheless, standard HLP models treat data as known and deterministic. This can result in highly sub-optimal solutions given the inherent uncertainty surrounding future conditions. There exist basically two streams of research dealing with optimization under uncertainty: stochastic optimization and robust optimization. In stochastic optimization, it is assumed that the values of the uncertain parameters are governed by known probability distributions. In robust optimization, it is assumed that parameters are uncertain but no information about their probability distributions is know except for the specification of intervals containing the uncertain values.

In classical facility location, stochastic models have been widely investigated over the last four decades. Louveaux (1986, 1993) presents classical reviews on modeling approaches for stochastic facility location in which the location of the facilities is considered as a first-stage decision and the distribution pattern is a second-stage decision. Some of these models (see Louveaux and Peeters, 1992; Laporte et al., 1994) consider capacities on the facilities, and facility size is considered as a first-stage decision. Ravi and Sinha (2006) propose a stochastic problem in which facilities may be open in either the first or second stage, while incurring different installation costs in each stage. The survey by Snyder (2006) covers both stochastic and robust location models for stochastic location problems.

To the best of the authors' knowledge, there exist only three published articles related to stochastic hub location problems. Marianov and Serra (2003) focus on stochasticity at the hub nodes by representing hub airports as $M/D/c$ queues and limiting through chance constraints the number of airplanes that can queue at an airport. The authors present a linear mixed integer programming (MIP) formulation and propose a heuristic procedure to obtain feasible solutions for instances with up to 30 nodes. Sim et al. (2009) introduce the stochastic p-hub center problem and employ a chance-constrained formulation to model the minimum service-level requirement. Their model takes into account the variability in travel times when designing the hub network so that the maximum travel time through the network is minimized. The authors present a linear MIP formulation for the problem, under the assumption that travel times on the arcs are independent normal random variables. They also propose several heuristics to obtain feasible solutions for instances with up to 25 nodes. Yang (2009) presents a stochastic model for air freight hub location and flight route planning under seasonal demand variations. The author models the problem as a two-stage stochastic program with recourse, in which the location of hub facilities is considered as a first-stage decision and the planning of flight routes as a second-stage decision. Stochasticity on demands as well as on discount factors on the arcs of the network is considered. Moreover, the model allows direct connections between non-hub nodes. A MIP formulation for the problem is presented under the assumption that demand is governed by a discrete probability distribution involving only three possible scenarios. A case study from a 10-node air freight market in Taiwan and China is also described.

One of the problems that have received most attention in deterministic hub location is the *Uncapacitated Hub Location Problem with Multiple Assignments* (UHLPMA). In this problem, the number of required hubs to locate is not known in advance, but a fixed set-up cost for each hub facility is considered. The capacity of the hubs and of the links of the hub network is unbounded. It is assumed that flows originating at the same node but having different destination points can be routed through different sets of hub nodes, i.e. a multiple assignment pattern applies. The objective is to minimize the sum of the hub fixed costs and demand routing costs. The best known formulations for the UHLPMA, in terms of LP relaxation bounds, are those of Hamacher et al. (2004) and of Marín et al. (2006), whereas the best exact algorithms are those of Cánovas et al. (2007), Camargo et al. (2008), and Contreras et al. (2010a). In particular, the Benders decomposition algorithm of Contreras et al. (2010a) can efficiently solve large-scale UHLPMA instances with up to 500 nodes.

In the UHLPMA, demand between O/D pairs as well as transportation costs are treated as known and deterministic. However, in real applications future demand is not known in advance and only a forecast may be available. Transportation costs between node pairs are usually defined to be proportional to the distance between nodes. However, transportation costs are also intimately related to the price of resources (fuel, electricity, raw materials) used to provide the actual transportation of demand, which may be highly uncertain. Other sources of uncertainty in transportation costs may be due to: *i)* uncertainty in travel distances, *ii)* traffic and congestion, *iii)* tariff changes by outsourcing companies, and *iv)* link failures. In this paper we study how the UHLPMA can be modeled as a two-stage integer stochastic program with recourse in the presence of uncertainty on demands and transportation costs. In particular, we introduce three different stochastic versions of the UHLPMA. The first is the *Uncapacitated Hub Location Problem with Stochastic Demands* (UHL-SD) in which demands between O/D pairs are considered to be stochastic. The second is the *Uncapacitated Hub Location Problem with Stochastic Dependent Transportation Costs* (UHL-SDC) where uncertainty is given by a single parameter influencing transportation costs. It is assumed that this parameter equally affects the transportation costs for all links of the network. The third is the *Uncapacitated Hub Location Problem with Stochastic Independent Transportation Costs* (UHL-SIC) in which the transportation costs are also stochastic. However, this problem considers the more general case in which the uncertainty of transportation costs is independent for each link of the network. We show that both UHL-SD and UHL-SDC are equivalent to their associated *Expected Value Problem* (EVP) in which uncertain transportation costs are replaced with their expected value (see Birge and Louveaux, 1997). However, this equivalence does not hold for the UHL-SIC.

We use a Monte-Carlo simulation-based method, known as the Sample Average Approximation (SAA) scheme (Kleywegt et al., 2001), to solve UHL-SIC problems with continuous distance distributions, and therefore, an infinite number of scenarios. This method can also be applied to UHL-SIC problems with a finite but very large number of scenarios. The idea of the SAA scheme is to generate a random sample and approximate the expected value function by the corresponding sample average function. The associated deterministic sample average optimization problem is then solved to obtain a solution of UHL-SIC, and the procedure is repeated. The SAA scheme not only generates high quality solutions when solving the sample average problems, but is also able to produce a statistical estimation of their optimality gap. We integrate a Benders decomposition scheme to solve the corresponding SAA problems.

The remainder of this paper is organized as follows. Sections 2 formally introduces two-stage stochastic models for the considered problems. Section 3 describes our solution method for the UHL-SIC. Computational results are presented in Section 4, followed by conclusions in Section 5.

2 Stochastic Uncapacitated Hub Location Problems

Before presenting the stochastic uncapacitated hub location models under study, we describe their deterministic counterpart, the UHLPMA. Let $G = (Q, A)$ be a complete digraph, where Q is the set of nodes and A is the set of arcs. Let also $H \subseteq Q$ represent the set of potential hub locations, and K be the set of commodities whose origin and destination points belong to Q. For each commodity $k \in K$, define W_k as the amount of commodity k to be routed from the origin $o(k) \in Q$ to the destination $d(k) \in Q$. For each node $i \in H$, f_i is the fixed set-up cost for locating a hub at node i. The transportation cost between nodes i and j is defined as $c_{ij} = \gamma d_{ij}$, where d_{ij} is the distance between nodes i and j, which is assumed to satisfy the triangle inequality, and γ is the resource cost per unit distance. All costs relate to the same planning horizon.

Given that hub nodes are fully interconnected and distances satisfy the triangle inequality, every path between an origin and a destination node will contain at least one and at most two hubs. For this reason, paths between two nodes are of the form $(o(k), i, j, d(k))$, where $(i, j) \in H \times H$ is the ordered pair of hubs to which $o(k)$ and $d(k)$ are allocated, respectively. Therefore, the unit transportation cost of routing commodity k along path $(o(k), i, j, d(k))$ is given by $F_{ijk} = \chi c_{o(k)i} + \tau c_{ij} + \delta c_{jd(k)}$, where χ, τ, and δ represent the collection, transfer and distribution costs along the path. To reflect economies of scale between hub nodes, we assume that $\tau < \chi$ and $\tau < \delta$. The UHLPMA consists in locating a set of hubs and in determining the routing of commodities through the hub nodes, with the objective of minimizing the total set-up and transportation cost.

We define binary location variables z_i, $i \in H$, equal to 1 if and only if a hub is located at node i. We also introduce binary routing variables x_{ijk}, $k \in K$ and $(i,j) \in H \times H$, equal to 1 if and only if commodity k transits via a first hub node i and a second hub node j. Following Hamacher et al. (2004), the UHLPMA can be stated as follows:

$$\text{minimize} \quad \sum_{i\in H} f_i z_i + \sum_{i\in H}\sum_{j\in H}\sum_{k\in K} W_k F_{ijk} x_{ijk} \tag{1}$$

$$\text{subject to} \quad \sum_{i\in H}\sum_{j\in H} x_{ijk} = 1 \qquad k \in K \tag{2}$$

$$\sum_{j\in H} x_{ijk} + \sum_{j\in H\setminus\{i\}} x_{jik} \le z_i \qquad i \in H, k \in K \tag{3}$$

$$x_{ijk} \ge 0 \qquad i,j \in H, k \in K \tag{4}$$

$$z \in \mathbb{B}^{|H|}. \tag{5}$$

The first term of the objective function represents the total set-up cost of the hub facilities and the second term is the total transportation cost. Constraints (2) guarantee that there is a single path connecting the origin and destination nodes of every commodity. Constraints (3) prohibit commodities from being routed via a non-hub node. Finally, constraints (4) and (5) are the standard non-negativity and integrality constraints. Given that there are no capacity constraints on the hub nodes, there is no need to explicitly state the integrality on the x_{ijk} variables because there always exists an optimal solution of (1)–(5) in which all x_{ijk} variables are integer.

We now present three stochastic variants of UHLPMA that incorporate uncertainty on demands W_k and transportation costs F_{ijk}. They can be formulated as two-stage stochastic programs with recourse, where the first-stage decisions correspond to the location of the hub facilities and the second-stage decisions correspond to the optimal routing of the commodities. As before, we use the z_i variables for the location of hubs and the x_{ijk} variables for the routing of commodities. However, given that the optimal routing of commodities depends on the particular realization of the random event ξ, we need to consider routing variables for each possible realization of ξ, i.e., $x_{ijk}(\xi)$, $k \in K$ and $(i,j) \in H \times H$, are binary routing variables equal to one if and only if commodity k transits via a first hub node i and a second hub node j for realization ξ.

2.1 Case A: Stochastic Demands

We consider the UHL–SD in which demand is uncertain. For each commodity $k \in K$, let $W_k(\xi)$ be a random variable representing the future demand of commodity k. It is assumed that the $W_k(\xi)$ are independent random variables, governed by known probability distributions, and that the distribution function of $\sum_{k\in K} W_k(\xi)$ can be computed. The

UHL–SD can be stated as:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i\in H} f_i z_i + E_\xi \left[\sum_{k\in K} \sum_{i\in H} \sum_{j\in H} \left(W_k(\xi) F_{ijk}\right) x_{ijk}(\xi) \right] && && (6)\\
\text{subject to} \quad & \sum_{i\in H} \sum_{j\in H} x_{ijk}(\xi) = 1 && k \in K,\ \xi \in \Xi && (7)\\
& \sum_{j\in H} x_{ijk}(\xi) + \sum_{j\in H\setminus\{i\}} x_{jik}(\xi) \le z_i && i \in H,\ k \in K,\ \xi \in \Xi && (8)\\
& x_{ijk}(\xi) \ge 0 && i, j \in H,\ k \in K,\ \xi \in \Xi && (9)\\
& z \in \mathbb{B}^{|H|}, && && (10)
\end{aligned}
$$

where E_ξ denotes the mathematical expectation with respect to ξ and Ξ is the support of ξ. The first term of the objective function represents the total set-up cost of the hub facilities and the second term is the total expected transportation cost with respect to ξ. Constraints (7)–(10) have the same meaning as in the case of UHLPMA, but they are defined for every possible realization of ξ.

We now show that the UHL–SD is in fact a deterministic problem in which the demands $W_k(\xi)$ can be replaced by their expectation.

Proposition 1 *The stochastic program UHL–SD is equivalent to the expected value problem:*

$$
\begin{aligned}
\textit{minimize} \quad & \sum_{i\in H} f_i z_i + \sum_{i\in H} \sum_{j\in H} \sum_{k\in K} \left(E_\xi \left[W_k(\xi)\right] F_{ijk}\right) x_{ijk} && (11)\\
\textit{subject to} \quad & (2) - (5).
\end{aligned}
$$

Proof Observe that, given a first-stage vector z, the second-stage term of the objective function can be separated into $|K|$ independent subproblems, one for each commodity $k \in K$. Furthermore, and more importantly, the optimal solution of each of these subproblems does not depend on the particular realization of the random variable ξ. That is, the optimal route for sending each commodity is the same regardless of the actual value of the demand $W_k(\xi)$. Let $x(z)$ be the optimal solution vector associated to a first-stage solution z. Then

$$
x_{ijk}(\xi) = x_{ijk}(z) \quad \forall k \in K,\ \forall i, j \in H,\ \forall \xi \in \Xi. \tag{12}
$$

By (12), the second-stage expectation of the objective function can be expressed as

$$
E_\xi \left[\sum_{i\in H} \sum_{j\in H} \sum_{k\in K} \left(W_k(\xi) F_{ijk}\right) x_{ijk}(z) \right] = \sum_{i\in H} \sum_{j\in H} \sum_{k\in K} \left(E_\xi \left[W_k(\xi)\right] F_{ijk}\right) x_{ijk}(z),
$$

where the equality follows from the fact that summations and expectation can be interchanged. The UHL-SD is therefore equivalent to the deterministic problem (11), (2)–(5). ■

2.2 Case B: Dependent Stochastic Transportation Costs

We now focus on the UHL–SDC in which uncertainty relates to the resource cost γ. Consequently, the transportation costs F_{ijk} are stochastic but dependent on a single uncertain parameter. In this problem the random variable γ is described by a known probability distribution. The unit transportation cost between every node pair (i,j) is $c^1_{ij}(\xi) = \gamma(\xi)d_{ij}$. Therefore, the unit transportation cost of routing commodity k along the path $(o(k), i, j, d(k))$ is given by $F^1_{ijk}(\xi) = \left(\chi c^1_{o(k)i}(\xi) + \tau c^1_{ij}(\xi) + \delta c^1_{jd(k)}(\xi)\right)$. The UHL–SDC can be stated as:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in H} f_i z_i + E_\xi \left[\sum_{k \in K} \sum_{i \in H} \sum_{j \in H} \left(W_k F^1_{ijk}(\xi) \right) x_{ijk}(\xi) \right] && (13) \\ \text{subject to} \quad & (7) - (10). \end{aligned}$$

We now show that the UHL–SDC can also be represented by a deterministic problem in which the transportation scale cost $\gamma(\xi)$ is replaced by its expectation.

Proposition 2 *The stochastic program UHL–SDC is equivalent to the expected value problem:*

$$\begin{aligned} \textit{minimize} \quad & \sum_{i \in H} f_i z_i + \sum_{i \in H} \sum_{j \in H} \sum_{k \in K} \left(W_k E_\xi \left[F^1_{ijk}(\xi) \right] \right) x_{ijk} && (14) \\ \textit{subject to} \quad & (2) - (5). \end{aligned}$$

Proof The proof uses the same idea as that of Proposition 1. Given a first-stage vector z, the second-stage term of the objective function can be separated into $|K|$ independent subproblems. Given that

$$\begin{aligned} F^1_{ijk}(\xi) &= \left(\chi\gamma(\xi) d_{o(k)i} + \tau\gamma(\xi) d_{ij} + \delta\gamma(\xi) d_{jd(k)} \right) \\ &= \gamma(\xi) \left(\chi d_{o(k)i} + \tau d_{ij} + \delta d_{jd(k)} \right), \end{aligned}$$

the optimal route for sending each commodity is the same regardless of the actual value of $\gamma(\xi)$. Let $x(z)$ be the optimal solution vector associated to a first-stage solution z. Then (12) follows and the second-stage expectation of the objective function can be expressed as

$$E_\xi \left[\sum_{i \in H} \sum_{j \in H} \sum_{k \in K} \left(W_k(\xi) F^1_{ijk} \right) x_{ijk}(z) \right] = \sum_{i \in H} \sum_{j \in H} \sum_{k \in K} \left(W_k E_\xi \left[F^1_{ijk}(\xi) \right] \right) x_{ijk}.$$

The UHL–SDC is thus equivalent to the deterministic problem (14), (2)–(5). ■

2.3 Case C: Independent Stochastic Transportation Costs

We now consider the UHL-SIC in which transportation costs F_{ijk} are stochastic. It is assumed that the random variables $d_{ij}(\xi)$, $i, j \in Q$, are independent and described by known probability distributions, and that for any $S \subseteq Q \times Q$, the distribution function of $\sum_{(i,j) \in S} d_{ij}(\xi)$

can be computed. In this case, transportation costs between nodes may no longer satisfy the triangle inequality. However, we assume that for each commodity k the path between origin and destination nodes will have at least one and at most two hub nodes. This assumption is often made even when transportation costs do not satisfy the triangle inequality (see, e.g., Marín et al., 2006; Contreras et al., 2010b). It is also consistent with practice. In ground transportation, for example, it is convenient to restrict each route to use at most two hub nodes so as to reduce handling and congestion in the hub facilities. In air transportation, it is also uncommon for a passenger to connect at more than two hub nodes.

The unit transportation cost between nodes i and j is $c^2_{ij}(\xi) = \gamma d_{ij}(\xi)$. Therefore, the unit cost of routing commodity k along the path $(o(k), i, j, d(k))$ is given by $F^2_{ijk}(\xi) = \left(\chi c^2_{o(k)i}(\xi) + \tau c^2_{ij}(\xi) + \delta c^2_{jd(k)}(\xi)\right)$. The UHL-SIC can be stated as

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in H} f_i z_i + E_\xi \left[\sum_{k \in K} \sum_{i \in H} \sum_{j \in H} \left(W_k F^2_{ijk}(\xi) \right) x_{ijk}(\xi) \right] && (15) \\ \text{subject to} \quad & (7) - (10). \end{aligned}$$

Observe that, in the case of the UHL-SIC, we cannot replace $F^2_{ijk}(\xi)$ variables by their expected value to obtain an equivalent deterministic problem because, contrary to the previous cases, the optimal solution of the second-stage depends on the particular realization of the random vector ξ.

We can extend some properties of optimal solutions of the UHLPMA to perform preprocessing to the UHL-SIC (see Contreras et al., 2010a). These properties lead to a more compact formulation with fewer variables, but with the same number of constraints. In particular, we define a set of candidate hub edges for each commodity $k \in K$ and for every possible realization of ξ as

$$E_k(\xi) = \begin{cases} \{e | e \in E, |e| = 1\} \bigcup A_k(\xi), & \text{if } o(k) \neq d(k), \\ \{e | e \in E, |e| = 1\}, & \text{otherwise,} \end{cases}$$

where

$$A_k(\xi) = \left\{ e : e \in E, |e| = 2 \text{ and } \left(F_{ek}(\xi) < \min \left\{ F_{\{e_1\}k}(\xi), F_{\{e_2\}k}(\xi) \right\} \right) \right\}.$$

The UHL–SIC can thus be restated as

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in H} f_i z_i + E_\xi \left[\sum_{k \in K} \sum_{e \in E_k(\xi)} \left(W_k F^2_{ek}(\xi) \right) x_{ek}(\xi) \right] && (16) \\ \text{subject to} \quad & \sum_{e \in E_k} x_{ek}(\xi) = 1 \qquad k \in K, \xi \in \Xi && (17) \\ & \sum_{e \in E_k(\xi): i \in e} x_{ek}(\xi) \leq z_i \qquad i \in H, k \in K, \xi \in \Xi && (18) \\ & x_{ek}(\xi) \geq 0 \qquad k \in K, \xi \in \Xi, e \in E_k(\xi) && (19) \\ & z \in \mathbb{B}^{|H|}. && (20) \end{aligned}$$

3 Algorithm for the UHL-SIC

We now present an algorithm for the UHL-SIC. The methodology incorporates a sampling technique, known as the SAA scheme (see Shapiro and Homem-De-Mello, 1998; Mak et al., 1999; Kleywegt et al., 2001), coupled with a Benders decomposition method (see Benders, 1962). The SAA scheme has previously been applied to obtain high quality solutions to stochastic supply chain design problems with a very large number of scenarios (Santoso et al., 2005; Schütz et al., 2009).

3.1 Sample Average Approximation

The main difficulty in solving the stochastic problem (15), (7)–(10) lies in the evaluation of the expected value of the objective function. In the SAA scheme, a random sample $N = \left\{\xi^i, \ldots, \xi^{|N|}\right\}$ of realizations of the random vector ξ is generated, and the second-stage expectation

$$E_\xi\left[\sum_{k\in K}\sum_{e\in E_k(\xi)}\left(W_k F_{ek}^2(\xi)\right)x_{ek}(\xi)\right]$$

is approximated by the sample average function

$$\frac{1}{|N|}\sum_{n\in N}\sum_{k\in K}\sum_{e\in E_{kn}}\widehat{F}_{ekn}x_{ekn},$$

where $\widehat{F}_{ekn} = W_k F_{ekn}^2$. Therefore, the original problem (15), (7)–(10) is approximated by the SAA problem

$$\begin{aligned}
\text{minimize}\quad & \sum_{i\in H} f_i z_i + \frac{1}{|N|}\sum_{n\in N}\sum_{k\in K}\sum_{e\in E_{kn}}\widehat{F}_{ekn}x_{ekn} && && (21)\\
\text{subject to}\quad & \sum_{e\in E_{kn}} x_{ek}^n = 1 && n\in N, k\in K && (22)\\
& \sum_{e\in E_{kn}:i\in e} x_{ekn} \le z_i && n\in N, i\in H,\ k\in K && (23)\\
& x_{ekn}\ge 0 && n\in N, k\in K, e\in E_{kn} && (24)\\
& z\in\mathbb{B}^{|H|}. && && (25)
\end{aligned}$$

Let v^N and $\hat{z}^N$ be the optimal solution value and an optimal solution, respectively, of the SAA problem (21)–(25). Observe that for a particular realization $\xi^i, \ldots, \xi^{|N|}$ of the random sample, the problem (21)–(25) is deterministic and can be solved by integer programming techniques.

It can be shown that under mild regularity conditions, v^N and $\hat{z}^N$ converge with probability one to their true counterparts, as the sample size $|N|$ increases, and furthermore $\hat{z}^N$ converges to an optimal solution of the true problem with probability approaching one exponentially fast (see Kleywegt et al., 2001). It is possible to estimate the sample size $|N|$

needed to generate an ε-optimal solution to the original problem, assuming that the SAA problem is solved within an absolute optimality gap $\delta \geq 0$, with probability at least equal to $1-\alpha$ whenever

$$|N| \geq \frac{3\sigma_{max}^2}{(\varepsilon-\delta)^2} \log\left(\frac{|\mathbb{B}^{|H|}|}{\alpha}\right), \tag{26}$$

where $\varepsilon > \delta$ and $\alpha \in (0,1)$. Here σ_{max}^2 is a maximal variance of certain function differences (see Kleywegt et al., 2001, for details). It is known that the sample size estimate (26) is too conservative for practical applications. However, one can choose a sample size $|N|$ as a trade-off between the quality of the associated optimal solution of the SAA problem (21)–(25) and the computational burden needed to solve it. Therefore, instead of solving one large-scale SAA problem, the SAA algorithm involves the repeated solution of the smaller SAA problem (21)–(25) with independent samples. Then, statistical confidence intervals are computed to evaluate the quality of the approximate solutions. We now describe this procedure.

1. Generate a set $M = \left\{N_1, \ldots, N_{|M|}\right\}$ of independent samples, each of size $|N|$, i.e., $\xi_j^i, \ldots, \xi_j^{|N|}$ for $j \in M$. For each sample N_j solve the corresponding SAA problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in H} f_i z_i + \frac{1}{|N|}\sum_{n\in N_j}\sum_{k\in K}\sum_{e\in E_{kn}} \widehat{F}_{ekn} x_{ekn} \\ \text{subject to} \quad & (22)-(25). \end{aligned} \tag{27}$$

   Let v^{N_j} and $\hat{z}^{N_j}$, $j \in M$, be the corresponding optimal objective value and an optimal solution, respectively.

2. Compute the average of all optimal solution values from the SAA problems and their variance:

$$\mu_M^N = \frac{1}{|M|}\sum_{j\in M} v^{N_j},$$

$$\sigma^2_{\mu_M^N} = \frac{1}{(|M|-1)|M|}\sum_{j\in M}\left(v^{N_j} - \mu_M^N\right)^2.$$

   It is known that the average μ_M^N provides a statistical lower bound for the optimal value of the original problem (15), (7)–(10) (Norkin et al., 1998; Mak et al., 1999), and $\sigma^2_{\mu_M^N}$ is an estimate of the variance of this estimator.

3. Choose a feasible solution $\hat{z} \in \mathbb{B}^{|H|}$ of the original problem, for instance, one of the previously obtained solutions $\hat{z}^{N_j}$. Using this solution, it is possible to estimate the optimal solution value v^* of the original problem (15), (7)–(10) as follows:

$$v_{N'}(\hat{z}) = \sum_{i\in H} f_i z_i + \frac{1}{|N'|}\sum_{n\in N'}\sum_{k\in K}\sum_{e\in E_{kn}} \widehat{F}_{ekn} x_{ekn},$$

   where $\xi^i, \ldots, \xi^{N'}$ is a sample of size N' generated independently of the samples employed in the SAA problems. Given that the first-stage variables are fixed, one can take $|N'|$ much larger than the sample size $|N|$ used in the SAA problems. Note that $v(\hat{z})$ is an estimate on the upper bound on the optimal solution value v^* of the original problem. The variance of this estimate can be computed as

$$\sigma^2_{N'}(\hat{z}) = \frac{1}{(|N'|-1)|N'|} \sum_{n \in N'} \left(\sum_{i \in H} f_i z_i + \sum_{k \in K} \sum_{e \in E_{kn}} \widehat{F}_{ekn} x_{ekn} - v(\hat{z}) \right)^2 .$$

4. Compute an estimate of the absolute optimality gap of solution $\hat{z}$ and its variance by using the lower and upper bound estimates on the optimal solution value of the original problem (15), (7)–(10) obtained in Steps 2 and 3.

$$\mathrm{gap}_{N,M,N'}(\hat{z}) = v_{N'}(\hat{z}) - \mu^N_M,$$
$$\sigma^2_{\mathrm{gap}} = \sigma^2_{\mu^N_M} + \sigma^2_{N'}(\hat{z}).$$

Using these estimators, we can construct a confidence interval for the optimality gap.

3.2 Benders Decomposition for the SAA Problem

The most computationaly expensive part of the SAA algorithm is the solution of $|M|$ two-stage stochastic integer problems (21)–(25) involving $|N|$ scenarios in Step 1. Even though problem (21)–(25) has far fewer scenarios than the original problem (15), (7)–(10), it is still a very large MIP problem that cannot be efficiently solved by means of a general-purpose solver, even for small size instances. We must therefore resort to a decomposition algorithm. In what follows, we present a Benders decomposition scheme for solving the SAA problem to optimality.

Benders decomposition is a well-known partitioning algorithm that separates an original mixed-integer problem into two simpler ones: an integer *master problem* and a linear *subproblem*. In this section, we introduce a Benders reformulation of the SAA problem and describe a Benders decomposition algorithm to solve it.

3.2.1 Benders Reformulation

For any fixed vector $\hat{z} \in \mathbb{B}^{|H|}$, the *primal subproblem* (PS) in the space of the x_{ekn} variables is

$$\begin{aligned} v(\hat{z}) = \text{minimize} \quad & \frac{1}{|N|} \sum_{n \in N} \sum_{k \in K} \sum_{e \in E_{kn}} \widehat{F}_{ekn} x_{ekn} \\ \text{subject to} \quad & (22), (25) \\ & \sum_{e \in E_{kn} : i \in e} x_{ekn} \le \hat{z}_i \qquad n \in N, i \in H, k \in K. \qquad (28) \end{aligned}$$

Let α_{kn} and u_{ikn} be the dual variables associated with constraints (22) and (28), respectively. The *dual subproblem* (DS), which is the dual of PS, can be stated as follows:

$$\begin{aligned}
\text{maximize} \quad & \sum_{k\in K}\sum_{n\in N}\alpha_{kn} - \sum_{i\in H}\sum_{k\in K}\sum_{n\in N}\hat{z}_i u_{ikn} && && (29)\\
\text{subject to} \quad & \alpha_{kn} - u_{e_1kn} - u_{e_2kn} \le \widehat{F}_{ekn} && n\in N, k\in K, e\in E_{kn}, |e|=2 && (30)\\
& \alpha_{kn} - u_{e_1kn} \le \widehat{F}_{ekn} && n\in N, k\in K, e\in E_{kn}, |e|=1 && (31)\\
& u_{ikn} \ge 0 && n\in N, k\in K, i\in H. && (32)
\end{aligned}$$

Let D denote the set of feasible solutions of DS and let P_D denote the set of extreme points of D. Observe that D is not modified when changing $\hat{z}$ and, because $F_{ekn} \ge 0$ for each $e \in E_k$ and $k \in K$, the null vector 0 is always a solution to DS. Hence, because of strong duality, either the primal subproblem is feasible and bounded, or it is infeasible. We are thus interested in $\hat{z}$ vectors that give rise to primal subproblems of the former case. It is known that if there exists at least one open hub facility, the primal and dual subproblems are always feasible and bounded (see Contreras et al., 2010a). Therefore, it follows that the dual objective function value is equal to

$$\max_{(\alpha,u)\in P_D}\sum_{k\in K}\sum_{n\in N}\alpha_{kn} - \sum_{i\in H}\sum_{k\in K}\sum_{n\in N}\hat{z}_i u_{ikn}. \qquad (33)$$

Introducing an extra variable η for the overall transportation cost, we can formulate the Benders master problem (MP) as follows:

$$\begin{aligned}
\text{minimize} \quad & \sum_{i\in H} f_i z_i + \eta && && \\
\text{subject to} \quad & \eta \ge \sum_{k\in K}\sum_{n\in N}\alpha_{kn} - \sum_{i\in H}\sum_{k\in K}\sum_{n\in N} u_{ikn} z_i && \forall(\alpha,u)\in P_D && (34)\\
& \sum_{i\in H} z_i \ge 1 && && (35)\\
& z \in \mathbb{B}^{|H|}. && && (36)
\end{aligned}$$

Observe that Benders feasibility cuts associated with the extreme rays of D are not necessary in the Benders reformulation because the feasibility of PS is ensured by constraints (35). We have thus transformed problem (21)–(25) into an equivalent MIP problem with $|H|$ binary variables and one continuous variable. Nevertheless, the above Benders reformulation contains an exponential number of constraints and must be tackled by an adequate cutting plane approach. Thus, we iteratively solve relaxed master problems containing a small subset of constraints (34) associated with the extreme points of P_D, and we keep adding these as needed by solving dual subproblems until an optimal solution to the original problem is obtained.

It is known that the number of cuts required to obtain an optimal solution can be reduced given that the subproblem is decomposable into $|K| \times |N|$ independent subproblems (see, e.g. Birge and Louveaux, 1988). We could in principle generate optimality cuts associated to extreme points of each dual polyhedron of the $|K|\times|N|$ subproblems. However, this approach may not be so effective given the large number of cuts to be added at each iteration. In fact, preliminary computational experiments have shown that the best Benders reformulation, in terms of required CPU times, is the one obtained by aggregating the information obtained to generate a set of optimality cuts associated with subsets of commodities. This reformulation has also been the most effective for solving the UHLPMA (Contreras et al., 2010a). In particular, for each node $i \in H$, let $K_i \subset K$ be the subset of commodities whose origin node is i. We can separate the subproblem into $|H|$ independent subproblems, one for each node. Hence, we consider the dual polyhedra of these $|H|$ subproblems and generate cuts from them. Let P_D be the set of extreme points of the dual polyhedron P_{D^i} associated with subproblem i. We thus obtain the following Benders reformulation:

$$\begin{aligned}
\text{minimize} \quad & \sum_{i\in H} f_i z_i + \sum_{i\in H} \eta_i \\
\text{subject to} \quad & (35), (36) \\
& \eta_i \geq \sum_{k\in K_i}\sum_{n\in N} \alpha^t_{kn} - \sum_{i\in H}\sum_{k\in K_i}\sum_{n\in N} u^t_{ikn} z_i \qquad \forall i \in H,\ (\alpha, u) \in P_{D^i}. \qquad (37)
\end{aligned}$$

Using this reformulation, only $|H|$ potential optimality cuts will be generated when solving the subproblem, instead of $|K| \times |N|$ cuts needed for the full separability into $|K| \times |N|$ dual subproblems.

3.2.2 Benders Decomposition Algorithm

Let ub denote an upper bound on the optimal solution value and let t represent the current iteration number. Let P^t_D denote the restricted set of extreme points of D at iteration t, MP(P^t_D) the relaxed master problem obtained by replacing P_D by P^t_D in MP, and $v(MP(P^t_D))$ its optimal solution value. Also, let z^t be an optimal solution vector of MP(P^t_D), DS(z^t) the dual subproblem for z^t, and $v(DS(z^t))$ its optimal solution value. A pseudo-code of the Benders decomposition algorithm is provided in Algorithm 1.

Algorithm 1: Benders decomposition

$ub \leftarrow \infty$, $t \leftarrow 0$
$P_D^t \leftarrow \varnothing$
$terminate \leftarrow$ **false**
while ($terminate =$ **false**) **do**
  Solve MP(P_D^t) to obtain z^t
  if ($v(MP(P_D^t)) = ub$) **then**
    $terminate \leftarrow$ **true**
  else
    Solve DS(z^t) to obtain $(\alpha^t, u^t) \in P_D$
    $P_D^{t+1} \leftarrow P_D^t \cup \{(\alpha^t, u^t)\}$
    if ($v(DS(z^t)) + \sum_{i \in H} f_i z_i^t < ub$) **then**
      $ub \leftarrow v(DS(z^t)) + \sum_{i \in H} f_i z_i^t$
    end if
  end if
  $t \leftarrow t + 1$
end while

Whenever the problem defined by (21)–(25) is feasible, Algorithm 1 will yield an optimal solution. We use the algorithm described in Contreras et al. (2010a) for efficiently solving DS(z^t) and obtain good optimality cuts at each iteration of the algorithm.

4 Computational Experiments

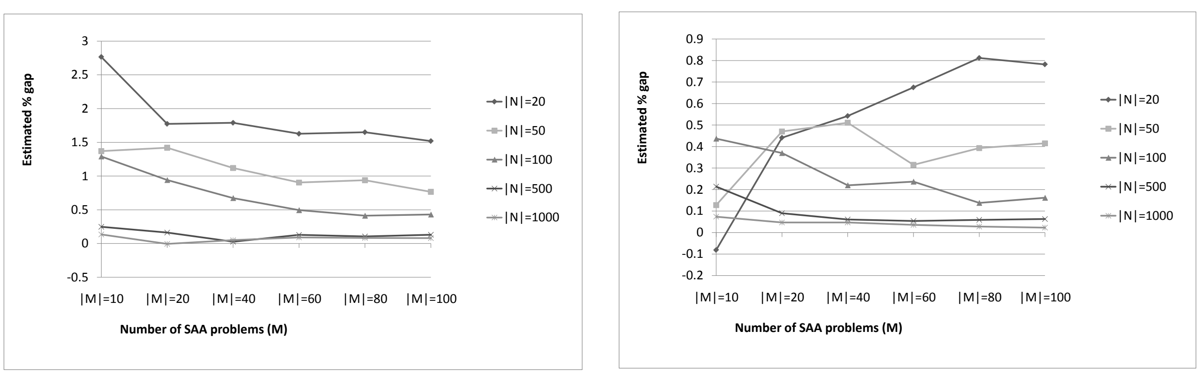
In this section we present the results of computational experiments performed to assess the behaviour of the proposed solution method. We first focus on some implementation details concerning the practical convergence of the SAA scheme, we then compare the effects of uncertainty under different continuous probability distributions and different levels of uncertainty on the optimal solutions of the UHL-SIC, and finally we test the robustness and limitations of our method on instances involving up to 50 nodes. All algorithms were coded in C and run on a Dell Studio PC with an Intel Core 2 Quad processor Q8200 running at 2.33 GHz and 8 GB of RAM under a Linux environment. The master problems of all versions of the algorithm to solve the SAA problems were solved using the callable library of CPLEX 10.1.

We have generated a set of benchmark instances using the procedure described in Contreras et al. (2010a). These instances consider different levels of magnitude for the amount of flow originating at a given node to obtain three different sets of nodes: low-level (LL) nodes, medium-level (ML) nodes, and high-level (HL) nodes. The total outgoing flow of LL, ML and HL nodes lies in the interval $[1, 10]$, $[10, 100]$, and $[100, 1000]$, respectively. Using these nodes, different classes of instances can be generated. We generate instances of the class *Set I* (see Contreras et al., 2010a, for details), in which the number of HL, ML, and LL nodes is 2%, 38% and 60% of the total number of nodes, respectively. We generate instances with $|H| = 10$, 20, 30, 40, and 50 and assume gamma and normal distributions for the uncertain transportation costs. We set the mean transportation costs c_{ij} between node i and j, for all $i, j \in H$, equal to the Euclidean distance and the standard deviation equal to $\sigma_{ij} = \nu \times c_{ij}$, where ν is the coefficient of variation. The non-negativity of the transportation costs is naturally preserved in the gamma distribution, whereas the normal distribution has to be truncated to avoid negative values.

4.1 Practical Convergence of SAA Algorithm

The aim of the first part of the computational experiments is to analyze the practical convergence of the SAA scheme. Recall that during the SAA procedure, we need to generate $|M|$ independent samples of size $|N|$. In order to select the proper sizes $|N|$ and $|M|$, some trade-offs need to be made. On the one hand, the objective function of the SAA problem tends to be a more accurate estimate of the objective function as we increase $|N|$. This implies that an optimal solution of the SAA problem tends to be a better solution and the corresponding estimate on the optimality gap tends to be tighter as $|N|$ increase. On the other hand, the computational complexity for the solution of the SAA problem increases at least linearly in $|N|$ and frequently exponentially. Therefore, it may be more efficient to choose a smaller sample of size $|N|$ and to increase $|M|$. We are interested in finding $|N|$ and $|M|$ such that the estimated optimality gap and the variance of the gap estimator are sufficiently small while keeping the required CPU time at minimum.

We test the practical convergence of the SAA algorithm by using different combinations of sample sizes $|N| \in \{20, 50, 100, 500, 1000\}$ and $|M| \in \{10, 20, 40, 60, 80, 100\}$. To perform this analysis, we select two 10-node instances of *Set I* having both $\tau = 0.2$ and $\nu = 0.5$, but with different probability distributions for the uncertain transportation costs. For each of the optimal solutions obtained in the SAA problems, we use a sample size of $|N'| = 100{,}000$ to obtain good estimations of the optimal solution value of the true problem (15), (7)–(10). Figures 1A and 1B plot the estimated optimality gap for different sample sizes $|N|$ and $|M|$ when considering the normal and gamma distributions for the uncertain transportation costs, respectively.

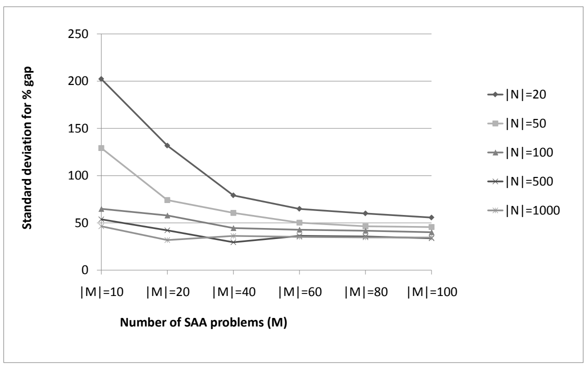


(A) Normal distribution (B) Gamma distribution

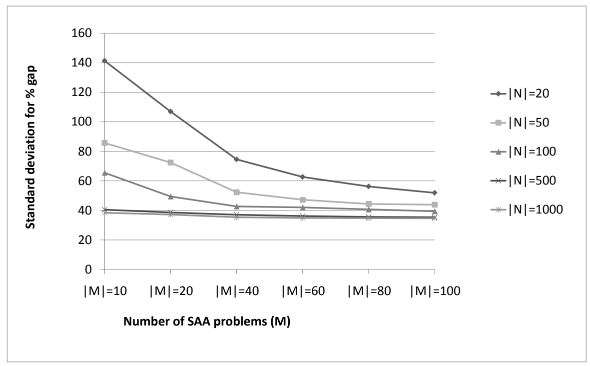
Figure 1: Optimality gap for a 10-node instance with different values of $|N|$ and $|M|$.

For the normal distribution (Figure 2A), we observe that by choosing a small sample size $|N| \leq 100$, the optimality gap remains above 0.5% even if we increase the sample size $|M|$. We can also appreciate that a higher number of SAA problems need to be solved in order to obtain a more accurate estimation of the optimality gap corresponding to a particular value of the sample size $|N|$. In contrast, when using a larger sample size such as $|N| = 500$ and $|N| = 1000$, only a small number of SAA problems are required to obtain more accurate optimality gaps below 0.1%. For the gamma distribution (Figure 2B), it is necessary to increase the number of SAA problems to solve in order to obtain accurate optimality gaps when using small samples $|N| \leq 100$. Observe that if $|N| = 20$, the estimated optimality gap is -0.1% if $|M| = 10$, whereas a more accurate estimation of 0.8% is obtained when increasing the number of SAA problems to $|M| = 100$. When using a larger sample $|N| \geq 500$, only a small number of SAA problems are needed to obtain an accurate optimality gap below 0.1%.

Figures 2A and 2B plot the estimated standard deviation for the optimality gap for different sample sizes of $|N|$ and $|M|$, with normal and gamma distributions, respectively.



(A) Normal distribution



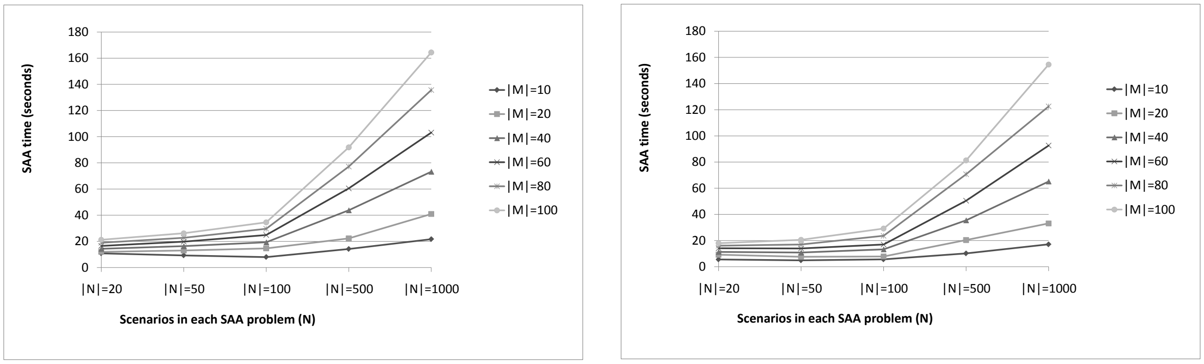
(B) Gamma distribution

Figure 2: Standard deviation for the optimality gap for a 10-node instance with different values of $|N|$ and $|M|$.

As expected, we observe that as the sample sizes $|N|$ and $|M|$ increase, the corresponding standard deviation for the optimality gap is considerably reduced for both the normal and gamma distributions. However, when considering large sample sizes $|N| \geq 500$ the standard deviation is sufficiently small, even if a small number of SAA problems are solved. In contrast, with sample sizes $|N| \leq 50$ a larger number of SAA problems are needed to reduce the variability.

Figures 3A and 3B plot the total CPU time required for the SAA algorithm for different sample sizes $|N|$ and $|M|$ with the normal and gamma distributions, respectively. From these figures, we observe that the computational complexity for solving the SAA problems seems to increase only linearly in $|N|$ in both cases. For instance, the required CPU time in the case of the normal distribution, $|N| = 500$ and $|M| = 80$ is 80 seconds whereas the CPU time increases to 140 seconds when $|N| = 1000$.

The results of the previous experiments indicate that it seems better to increase the sample size $|N|$ rather than the number of SAA problems when applying the SAA algorithm to the UHL-SIC. Furthermore, our results also indicate that when using a large sample size



(A) Normal distribution (B) Gamma distribution

Figure 3: Total CPU time for a 10-node instance with different values of $|N|$ and $|M|$.

$|N|$, only a small number of SAA problems are required to obtain tight optimality gaps having a small deviation. For these reasons, during the rest of the computational experiments we use sample sizes $|N| = 1000$ and $|M| = 20$.

4.2 Effects of Uncertainty on Optimal Solutions

The second part of the experiments is devoted to studying the effects of introducing uncertainty in the transportation costs of the UHLPMA. We use the coefficient of variation as a control parameter to introduce different levels of uncertainty into the model. In particular, we consider $\nu \in \{0.25, 0.5, 0.75, 1.0, 2.0\}$ to represent a wide range of situations for the amount of uncertainty in the transportation costs. We compare the stochastic solutions of UHL-SIC obtained with the SAA algorithm to those of the optimal solution of EVP.

We first study the effects of uncertainty by using the two 10-node instances of *Set I* used in the previous experiments. The computational results for the normal and gamma distributions are summarized in Tables 1 and 2, respectively. The first column provides the coefficient of variation ν whereas the next two columns give the estimated objective value relative to the EVP and the best solution provided by the SAA algorithm, respectively. The *Lower bound SAA* column provides the statistical lower bound obtained with the SAA algorithm, i.e. the average μ_M^N. The next two columns under the heading *Estimated % gap* provide the percent optimality gap relative to the EVP and the best solution provided by the SAA algorithm, respectively. We recall that the gaps are computed with respect to the lower bound obtained by the SAA algorithm. The next column gives the 95% confidence interval for the optimality gap of the best solution obtained by the SAA algorithm. We assume that $\text{gap}_{N,M,N'}(\hat{z})$ is normally distributed with variance σ^2_{gap}. The next column under the heading *Time (sec)* gives the total time in seconds for required for the SAA algorithm. The last column provides the best found solution for the SAA algorithm. The first row with $\nu = 0.00$ corresponds to the optimal EVP solution.

The results presented in Table 1 show that, for the considered instance, the optimal EVP solution does not remain optimal when uncertainty is introduced in the transportation costs. The SAA algorithm yields in all cases a better solution having a smaller optimality gap than

Table 1: Comparison for a 10-node instance of *Set I* using a normal distribution.

ν	*Objective value* *EVP*	*SAA*	*Lower bound* *SAA*	*Optimality % gap* *EVP*	*SAA*	*CI for SAA* *% gap at 95%*	*Time* *(sec)*	*Best* *solution*
0.00	25483.75	-	-	-	-	-	0.14	1,5,8
0.25	25333.91	25224.59	25214.35	0.47	0.04	(-0.13,0.21)	28.68	1,5
0.50	24976.67	24255.89	24244.58	2.93	0.05	(-0.21,0.30)	34.29	5,7,9
0.75	25182.79	23859.95	23860.72	5,25	0.00	(-0.27,0.26)	38.56	5,7,9
1.00	25973.57	24344.77	24276.20	6.53	0.28	(0.00,0.56)	56.41	5,7,9
2.00	31434.09	28188.67	28161.16	10.41	0.10	(-0.21,0.41)	76.38	4,5,7,9

that of the EVP. Moreover, this difference becomes more important when the variability in the uncertain transportation costs increases. For a low variability level $\nu = 0.25$, the optimal (or best known) set of hubs is reduced to only two hubs by eliminating hub node 8. However, for medium variability levels $\nu = 0.50$, 0.75 and 1.0, the set changes to be the hub nodes 5, 7 and 9. In the case of a high variability level $\nu = 2.0$, the cardinality of the set increases by one. Note that node 5 is the only hub node that appears in all considered cases of uncertainty.

Table 2: Comparison for a 10-node instance of *Set I* using a gamma distribution.

ν	*Objective value* *EVP*	*SAA*	*Lower bound* *SAA*	*Optimality % gap* *EVP*	*SAA*	*CI for SAA* *% gap at 95%*	*Time* *(sec)*	*Best* *solution*
0.00	25483.75	-	-	-	-	-	0.14	1,5,8
0.25	24543.10	24152.47	24130.73	1.68	0.09	(-0.12,0.30)	34.94	1,5
0.50	23479.64	22851.23	22852.34	2.67	0.00	(-0.23,0.22)	39.76	1,5,7
0.75	22573.51	21595.95	21598.96	4.32	-0.01	(-0.29,0.26)	42.01	1,5,7
1.00	21795.88	20584.52	20564.26	5.65	0.10	(-0.20,0.39)	47.05	1,5,7
2.00	19525.83	17846.37	17827.08	8.70	0.11	(-0.23,0.45)	60.35	1,5,7

Similar observations can be drawn from Table 2 for the gamma distribution. When $\nu = 0.25$, the best set of hub nodes is the same as for the normal distribution. However, for higher levels of variability $\nu \geq 0.5$, the set of hubs becomes $\{1, 5, 7\}$. Note that nodes 1 and 5 are always selected as hub nodes in all considered cases of uncertainty.

We have also used a 10-node instance from the well-known AP (Australian Post) set of instances to study the effects of uncertainty. This data set is the most commonly used in the hub location literature (*mscmga.ms.ic.ac.uk/jeb/orlib/phubinfo.html*). Transportation costs are proportional to the Euclidean distances e_{ij} between 200 cities in Australia and the values of W_k represent postal flows between pairs of cities. From this set of instances, we select the 10-node instances with set-up costs of the type loose (L), inter-hub discount cost $\tau = 0.2$, and compute the expected transportation cost as $d_{ij} = TC \times e_{ij}$ for each pair $(i, j) \in H \times H$, where $TC = 3$ is a scaling parameter for the transportation costs (see Contreras et al., 2010a, for details). The computational results for the normal and gamma distributions are summarized in Tables 3 and 4, respectively.

The results presented in Table 3 show that the optimal EVP solution remains the best (provably optimal) solution when uncertainty is introduced in the transportation costs, even for the largest considered level of variability $\nu = 2.0$. Similar results can be drawn from Table 4 for the gamma distribution. Only for the highest level of variability $\nu = 2.0$, is the SAA algorithm capable of finding a slightly better solution than the optimal EVP solution. Additional computational experiment have shown that similar results are obtained for the majority of AP data instances with up to 50 nodes. This situation may be partially

Table 3: Comparison for a 10-node instance of AP using a normal distribution.

	Objective value		*Lower bound*	*Optimality % gap*		*CI for SAA*	*Time*	*Best*
ν	*EVP*	*SAA*	*SAA*	*EVP*	*SAA*	*% gap at 95%*	*(sec)*	*solution*
0.00	189568.81	-	-	-	-	-	0.12	1,4,7
0.25	189503.05	189503.05	189498.61	0.00	0.00	(-0.24,0.24)	27.82	1,4,7
0.50	189427.06	189427.06	189423.11	0.00	0.00	(-0.24,0.24)	28.01	1,4,7
0.75	189332.76	189332.76	189342.71	-0.01	-0.01	(-0.25,0.23)	29.11	1,4,7
1.00	189222.63	189222.63	189216.51	0.00	0.00	(-0.24,0.24)	28.65	1,4,7
2.00	188847.70	188847.70	188844.77	0.00	0.00	(-0.24,0.24)	28.33	1,4,7

Table 4: Comparison for a 10-node instance of AP using a gamma distribution.

	Objective value		*Lower bound*	*Optimality % gap*		*CI for SAA*	*Time*	*Best*
ν	*EVP*	*SAA*	*SAA*	*EVP*	*SAA*	*% gap at 95%*	*(sec)*	*solution*
0.00	189568.81	-	-	-	-	-	0.12	1,4,7
0.25	189497.02	189497.02	189491.78	0.00	0.00	(-0.24,0.24)	27.62	1,4,7
0.50	189128.30	189128.30	189118.00	0.01	0.01	(-0.24,0.25)	27.70	1,4,7
0.75	188511.60	188511.60	188493.37	0.01	0.01	(-0.25,0.24)	29.74	1,4,7
1.00	187621.25	187621.25	187584.65	0.02	0.02	(-0.22,0.26)	32.51	1,4,7
2.00	183165.52	183085.70	182979.98	0.10	0.06	(-0.12,0.24)	46.57	4,7

explained by the fact that the AP instances are known to have a very peculiar flow structure. In particular, it is known that the amount of flow originating at each node is highly variable in every instance of this set: all instances have a very small number of nodes for which the outgoing flow is much larger than for the other nodes (see Contreras et al., 2010a). This situation seems to bias the optimal location of the hubs since very few nodes have a large impact on the overall cost of the network and thus greatly influence the hub location decisions. Therefore, even if we introduce a high variability in the uncertain transportation costs, the optimal EVP solution is still the best one.

4.3 Solving Medium-Size Instances

In the last part of the computational experiments, we analyze and evaluate the performance of the SAA algorithm on medium-size instances with up to 50 nodes. The computational results for the normal and gamma distributions are summarized in Tables 5 and 6, respectively. The first three columns provide the number of nodes, the discount factor and the coefficient of variation. The other columns have the same meaning as in the previous tables.

The results of Table 5 confirm the efficiency of the SAA algorithm for the UHL-SIC. The estimated optimality gaps of the best solutions provided by the SAA algorithm are always below 0.1%, except for one instance with 0.28%, and the upper bounds of the confidence intervals never exceed 0.56%. Moreover, the SAA algorithm is able to improve the solution of the expected value problem in 19 out of the 20 tested instances. Nevertheless, the CPU time required for the SAA algorithm grows very fast when instance size and variability increase.

Similar results can be drawn from Table 6 for the gamma distribution. The estimated optimality gap provided by the SAA algorithm is always below 0.1% and the upper bound for the confidence interval never exceeds 0.46%.

Table 5: Computational results for 20 instances of *Set I* using a normal distribution.

Instance			*Objective value*		*LB*	*% Gap*		*CI for SAA*	*Time*	*Best solution*	
$\|H\|$	τ	ν	*EVP*	*SAA*	*SAA*	*EVP*	*SAA*	*% gap*	*SAA*	*EVP*	*SAA*
10	0.2	0.5	24976.67	24255.89	24244.58	2.93	0.05	(-0,21,0,30)	34.29	1,5,8	5,7,9
10	0.7	0.5	28220.7	28220.70	28198.95	0.08	0.08	(-0,12,0,28)	19.49	1,5	1,5
10	0.2	1.0	25973.57	24344.77	24276.20	6.53	0.28	(0,00,0,56)	56.41	1,5,8	5,7,9
10	0.7	1.0	29650.95	29355.77	29325.46	1.10	0.10	(-0,15,0,36)	37.22	1,5	1,5,9
20	0.2	0.5	41060.47	40757.68	40751.59	0.75	0.01	(-0,21,0,24)	791.07	2,20	2,4,20
20	0.7	0.5	48853.55	48695.73	48693.27	0.33	0.01	(-0,15,0,16)	374.20	2,20	2,11
20	0.2	1.0	44758.35	41915.99	41877.31	6.44	0.09	(-0,14,0,33)	1740.00	2,20	2,11,20
20	0.7	1.0	53093.10	50712.01	50752.43	4.41	-0.08	(-0,38,0,18)	1499.68	2,20	2,11,19,20
30	0.2	0.5	149425.70	148219.39	148219.39	0.81	0.00	(-0,20,0,20)	5332.18	7,24,28	6,7,24
30	0.7	0.5	178453.27	175772.29	175765.89	1.51	0.00	(-0,16,0,17)	2421.57	7,18	6,7,17
30	0.2	1.0	151485.62	148820.66	148849.47	1.74	-0.02	(-0,26,0,22)	18329.18	7,24,28	7,11,17,24
30	0.7	1.0	184972.61	173460.65	173356.94	6.28	0.06	(-0,14,0,26)	13759.99	7,18	1,7,17
40	0.2	0.5	112094.75	112094.75	112129.65	-0.03	-0.03	(-0,28,0,18)	11582.06	9,17,20	9,17,20
40	0.7	0.5	129931.89	125250.34	125185.79	3.65	0.05	(-0,09,0,19)	6169.20	9,21	9,28
40	0.2	1.0	121937.39	120169.73	120208.65	1.42	-0.03	(-0.34,0.27)	52745.89	9,17,20	9,20,28
40	0.7	1.0	146287.70	135070.60	135129.94	7.63	-0.04	(-0,26,0,17)	30930.29	9,21	9,28,34
50	0.2	0.5	65622.75	59880.41	59925.68	8.68	-0.08	(-0,39,0,23)	23458.24	7	43,50
50	0.7	0.5	65635.07	62501.18	62562.61	4.68	-0.10	(-0,33,0,09)	14563.79	7	43,50
50	0.2	1.0	90642.87	72518.57	72556.89	19.95	-0.05	(-0,47,0,28)	86400.00	7	43,50
50	0.7	1.0	95781.32	76082.17	76141.13	20.51	-0.08	(-0,55,0,22)	86400.00	7	43,50

5 Conclusions

We have introduced stochastic uncapacitated hub location problems in which demands or transportation costs are uncertain. We have proved that stochastic problems with uncertain demands or dependent transportation costs are equivalent to a deterministic expected value problem. For the uncertain independent transportation costs, we have presented a solution method that integrates a sampling strategy, the SAA scheme, coupled with a Benders decomposition algorithm to obtain solutions to problems with a very large number of scenarios. Computational results confirm the efficiency and robustness of the proposed solution method. Benchmark instances involving up to 50 nodes were solved within an estimated optimality gap inferior to 0.3%. Furthermore, we have shown that the stochastic solutions obtained with the SAA algorithm are superior to those of the expected value problem, and the relative difference between the solution costs of the two problems tends to increase with the variance of the uncertain transportation costs.

Acknowledgments

This work was partly funded by the Canadian Natural Sciences and Engineering Research Council under grants 227837-09 and 39682-10. This support is gratefully acknowledged.

References

S. Alumur, B.Y. Kara, Network hub location problems: The state of the art, European Journal of Operational Research, 190 (2008) 1–21.

J.F. Benders, Partitioning procedures for solving mixed variables programming problems, Numersiche Mathematik, 4 (1962) 238–252.

Table 6: Computational results for 20 instances of *Set I* using a gamma distribution.

Instance			*Objective value*		*LB*	*% Gap*		*CI for SAA*	*Time*	*Best solution*	
$\lvert H\rvert$	τ	ν	*EVP*	*SAA*	*SAA*	*EVP*	*SAA*	*% gap*	*SAA*	*EVP*	*SAA*
10	0.2	0.5	23479.64	22851.23	22852.34	2.67	0.00	(-0.23,0.22)	40.35	1,5,8	1,5,7
10	0.7	0.5	26434.96	26434.96	26459.96	-0.09	-0.09	(-0.36,0.11)	26.28	1,5	1,5
10	0.2	1.0	21795.88	20584.52	20564.26	5.65	0.10	(-0.20,0.39)	46.17	1,5,8	1,5,7
10	0.7	1.0	24197.02	23963.12	23988.48	0.86	-0.11	(-0.39,0.18)	41.88	1,5	1,5,9
20	0.2	0.5	37999.81	37026.13	37011.09	2.60	0.04	(-0.25,0.33)	1317.00	2,20	2,11,17,20
20	0.7	0.5	45747.11	44204.07	44171.33	3.44	0.07	(-0.14,0.29)	629.02	2,20	2,11,19
20	0.2	1.0	34784.15	32720.74	32704.24	5.98	0.05	(-0.24,0.34)	1976.93	2,20	2,19,20
20	0.7	1.0	42144.67	38172.47	38210.97	9.33	-0.10	(-0.37,0.17)	1532.53	2,20	2,11,19
30	0.2	0.5	138954.48	137273.53	137372.54	1.14	-0.07	(-0.29,0.14)	8755.38	7,24,28	6,7,24
30	0.7	0.5	168325.23	160425.78	160469.86	4.67	-0.03	(-0.22,0.17)	5644.40	7,18	7,17,18
30	0.2	1.0	125171.77	122668.58	122737.18	1.94	-0.06	(-0.29,0.14)	20120.32	7,24,28	7,17,24,28
30	0.7	1.0	154033.43	136881.10	136962.42	11.08	-0.06	(-0.32,0.20)	14811.53	7,18	7,28,34
40	0.2	0.5	97487.40	96272.78	96239.99	1.28	0.03	(-0.24,0.31)	26837.41	9,17,20	7,9,28
40	0.7	0.5	129931.89	125250.34	125185.79	3.65	0.05	(-0.09,0.19)	6212.16	9,21	9,28
40	0.2	1.0	86201.89	82950.37	82876.56	3.86	0.09	(-0.26,0.24)	38808.13	9,17,20	7,9,28
40	0.7	1.0	105397.52	87696.63	87772.61	16.72	-0.09	(-0.54,0.46)	42583.25	9,21	7,28,34
50	0.2	0.5	58777.56	43532.81	43550.68	25.91	-0.04	(-0.37,0.28)	26863.32	7	43,50
50	0.7	0.5	58797.57	45465.50	45480.25	22.25	-0.03	(-0.34,0.28)	34296.60	7	43,50
50	0.2	1.0	58737.43	37982.81	38012.08	35.28	-0.08	(-0.54,0.23)	25204.74	7	43,50
50	0.7	1.0	58784.56	39241.18	39269.24	33.20	-0.07	(-0.48,0.27)	29628.68	7	43,50

J.R. Birge, F.V. Louveaux, Introduction to stochastic programming, Springer-Verlag New York, 1997.

J.R. Birge, F.V. Louveaux. A multicut algorithm for two-stage stochastic linear programs, European Journal of Operational Research 34 (1988) 384–392.

R.S. Camargo, G. Miranda Jr., H.P. Luna. Benders decomposition for the uncapacitated multiple allocation hub location problem, Computers & Operations Research 35 (2008) 1047–1064.

J.F. Campbell, A.T. Ernst, M. Krishnamoorthy, Hub location problems, in: Z. Drezner, H.W. Hamacher (Eds.), Facility Location: Applications and Theory, Springer, Heidelberg, 2002, pp. 373–408.

L. Cánovas, S. Garcia, A. Marín. Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique, European Journal of Operational Research, 179 (2007) 990–1007.

I. Contreras, J.-F. Cordeau, G. Laporte, Benders decomposition for large-scale uncapacitated hub location, Techincal Report CIRRELT-2010-26, HEC Montréal (2010a).

I. Contreras, E. Fernádez, A. Marín, The tree of hubs location problem, European Journal of Operational Research, 202 (2010b) 390–400.

H.W. Hamacher, M. Labbé, S. Nickel, T. Sonneborn, Adapting polyhedral properties from facility to hub location problems, Discrete Applied Mathematics, 145 (2004) 104–116.

A. J. Kleywegt, A. Shapiro, T. Homem-De-Mello, The sample average approximation method for stochastic discrete optimization, SIAM Journal of Optimization, 12 (2001) 479–502.

G. Laporte, F.V. Louveaux, L. Van hamme, Exact solution to a location problem with stochastic demands, Transportation Science, 28 (1994) 95–103.

F.V. Louveaux, Discrete stochastic location models, Annals of Operations Research, 6 (1986) 23–34.

F.V. Louveaux, Stochastic location analysis, Location Science 1 (1993) 127–154.

F.V. Louveaux, D. Peeters, A dual-based procedure for stochastic facility location, Operations Research, 40 (1992) 564–573.

W.K. Mak, D.P. Morton, R.K. Wood, Monte-Carlo bounding techniques for determining solution quality in stochastic programs, Operations Research Letters, 24 (1999) 47–56.

V. Marianov, D. Serra, Location models for airline hubs behaving as M/D/c queues, Computers & Operations Research, 30 (2003) 983–1003.

A. Marín, L. Canovas, M. Landete. New formulations for the uncapacitated multiple allocation hub location problem, European Journal of Operational Research, 172 (2006) 274–292.

V.I. Norkin, G.C. Pflug, A. Ruszczynski, A branch and bound method for stochastic global optimization, Mathematical Programming, 83 (1998) 425–450.

R. Ravi, A. Sinha, Hedging uncertainty: Approximation algorithms for stochastic optimization problems, Mathematical Programming Series A, 108 (2006) 97–114.

T. Santoso, S. Ahmed, M. Goetschalckx, A. Shapiro, A stochastic programming approach for supply chain network design under uncertainty, European Journal of Operational Research, 167 (2005) 96–115.

P. Schütz, A. Tomasgard, S. Ahmed, Supply chain design under uncertainty using sample average approximation and dual decomposition, European Journal of Operational Research, 199 (2009) 409–419.

A. Shapiro, T. Homem-De-Mello, A simulation-based approach to stochastic programming with recourse, Mathematical Programming 81 (1998) 301–325.

T. Sim, T.J. Lowe, B.W. Thomas, The stochastic p-hub center problem with service-level constraints, Computers & Operations Research, 36 (2009), 3166–3177.

L.V. Snyder, Facility location under uncertainty: A review, IIE Transactions 38 (2006) 537–554.

L.V. Snyder, M.S. Daskin, Stochastic p-robust location problems, IIE Transactions 38 (2006) 971–985.

T.-H. Yang, Stochastic air freight hub location and flight routes planning, Applied Mathematical Modeling, 33(12) (2009) 4424–4430.