

An alternative approach to address uncertainty in hub location

Marc Janschekowitz^a, Gita Taherkhani^b, Sibel A. Alumur^c, Stefan Nickel^a

*marc.janschekowitz@ymail.com, gtaherkhani@luc.edu, sibel.alumur@uwaterloo.ca,
stefan.nickel@kit.edu*

^a Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

^bQuinlan School of Business, Loyola University, Chicago, Illinois, USA

^cDepartment of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada

Abstract

In this paper, optimization and simulation techniques are integrated to address single and multiple allocation hub network design problems under uncertainty. Using a scenario-based iterative optimization-simulation approach four sources of uncertainty are considered: the demand to be transported within the network, the associated transportation costs as well as the fixed costs for both opening hub facilities and establishing the connections between them. Additionally, flow-dependent economies of scale on all network connections are incorporated in the simulation phase. A value of simulation measure is introduced to evaluate the performance of the methodology. The computational tests conducted on the well-known CAB data set with varying levels of uncertainty show that the approach can result in better solutions with up to 6.6% lower cost compared to its deterministic counterpart.

Keywords: Location, hub network design, iterated optimization-simulation.

1 Introduction

Having first been introduced by O’Kelly [38], *hub location problems* have since then become an important and intensely researched field of location theory. Telecommunication systems, computer networks and modern transportation, especially passenger air transport and postal delivery, all depend on the ability to efficiently route passengers and flows between a large number of origins and destinations within a network. Although many extensions regarding applicability and realism have been made to the original hub location model, the basic concept still remains unchanged: instead of directly serving each origin-destination pair in a many-to-many network, *hub facilities* are used to concentrate flows on few inter-hub connections. This allows for savings

to be established due to a smaller number of links and the exploitation of economies of scale. In this context, hubs function as switching, sorting, connecting and consolidation points in the network.

The fundamental decisions in hub location problems are to address where to locate hub facilities and how to design the hub network for routing the demand from its origin to the destination via hub facilities. Depending on the specific problem case, non-hub nodes are either assigned to exactly one (single allocation) or at least one (multiple allocation) hub facility. One of the most commonly used objectives is to minimize total costs caused by establishing and using this hub network in order to satisfy the demand of all origin-destination pairs.

The location of hubs is generally considered to be part of a strategic decision-making process due to its importance for the overall performance of a network for a long planning horizon. Hence, not only the operational short-term effects, but also long-term consequences have to be considered. A long-term planning horizon, however, makes anticipating future events and the corresponding data challenging, as a multitude of possible influences and changes may happen during this time period. Consequently, data can hardly be assumed to be deterministic and uncertainty has to be incorporated in the planning process. Neglecting the influence of uncertainty on parameters like demand, transportation costs, or set-up costs for hubs and hub arcs, which for example may vary due to inflation, fluctuating prices of property and raw materials, or unforeseen events such as COVID, can be especially problematic because hub location decisions may be linked to cost-intensive projects due to their strategic importance. For example, for the design of express shipment networks, considerable investments need to be made for locating sorting centers to act as hub facilities.

This paper presents a methodology to address uncertainty in hub location problems by using an iterative scenario-based optimization-simulation approach. First, single and multiple allocation versions of the hub network design problem incurring decisions on the establishment of inter-hub connections are formulated. Then, uncertainty is incorporated into the problem through a combination of optimization and simulation techniques in which an optimization model is iteratively solved to find a “good” candidate solution, while a simulation is used as an evaluation method for the solution obtained from this optimization model. The overall goal is to obtain solutions that perform well under uncertainty. Hence, several random scenarios are generated in the simulation phase based on various sources of uncertainty. Fluctuation in demand, in transportation costs, as well as in fixed costs for both opening hub facilities and operating inter-hub links are considered. Additionally, flow-dependent economies of scale on all network connections and different levels of uncertainty are modeled in the simulation. A metric, referred to as the *value of simulation* is formulated to measure the performance of the

methodology.

We would like to remark that this proposed iterative optimization-simulation approach is different than the classical simulation-based optimization (or simulation-optimization) technique where optimization is used as a tool for determining the values of the input variables for simulation (for more information on simulation-optimization see, e.g., Carson and Maria [17] and Nguyen et al. [36]). Our iterative optimization-simulated method can be considered more similar in nature to two-stage stochastic optimization. However, the second stage of our approach (i.e., the simulation phase) needs less information on the uncertain parameters and is also applicable to cases where the distribution of the uncertain parameters is not known or available. Moreover, the use of an iterated optimization-simulation approach can be more convenient and flexible than stochastic optimization for incorporating various types of uncertainty especially when there is a stochastic dependency between the problem parameters or when the distribution of the uncertain parameters is unknown (see, e.g., Grigoryev [26], Heckmann [27], and Bakker et al. [8]).

The models and methods that we develop in this paper can be used to derive insights into optimal hub locations for various practical applications that inherit uncertainty. Even though we take uncertainty in several problem parameters into account, we do not consider a dynamic planning horizon. As noted in Alumur et al. [4], incorporating the time dimension is also important in the design of hub networks. In this paper, however, we model a single period setting where all values of the parameters are scaled considering the length of this decision period. Origin-destination demands can be considered as a rate in units of flow per one period of time and this rate is subject to uncertainty. Express shipment network design can be an example of such an application where the demand can be expressed as a rate in units of flow, for example, per day or week.

Our contributions in this paper are (i) proposing an iterative optimization-simulation methodology to address uncertainty in hub location problems, (ii) implementing this methodology for single and multiple allocation hub network design formulations, (iii) presenting computational analysis under several sources of uncertainty using a benchmark data set, (iv) demonstrating that the proposed method can result in better solutions than deterministic optimization under uncertainty. This paper can be considered a “proof-of-concept” for further and more complex joint implementations of optimization and simulation to address uncertainty in hub location problems.

The remainder of this paper is organized as follows: In the next section, the related literature is reviewed. In Section 3, the mathematical formulations of the deterministic setting are presented. In Section 4, the generalized hybrid solution methodology is introduced. Section 5 includes the performed computational tests and the evaluation of the obtained solutions. Fi-

nally, Section 6 provides concluding remarks and a short discussion on possible future research directions.

2 Literature Review

Hub location problems presented in the literature are commonly classified by their allocation strategy referred to as single and multiple allocation. In the first setting, every non-hub node is allocated to exactly one hub (e.g., Ernst and Krishnamoorthy [23], O’Kelly [39], and Yaman [54]). In the latter setting, non-hub nodes can be allocated to more than one hub node (e.g., Campbell [12], Ernst and Krishnamoorthy [24], and Boland et al. [10]). Some papers considered and modeled both versions, e.g., Campbell [13], Skorin-Kapov et al. [50], and O’Kelly et al. [40].

The earlier hub location research focused on building fully interconnected inter-hub networks (e.g., O’Kelly [38, 39], Campbell [13], Ernst and Krishnamoorthy [23], and Skorin-Kapov et al. [50]). For many studies, this is a direct consequence of having triangle inequality for transportation costs and not having any fixed costs for inter-hub links in the models. More recently, inter-hub network design decisions are incorporated into hub location models resulting in more realistic formulations of this real-life problem (e.g., Nickel et al. [37], Campbell et al. [14], Alumur et al. [5], Alumur et al. [6], and Taherkhani and Alumur [51]).

In most hub location models, the economies of scale is modeled by a constant discount factor, referred to as $\alpha \in [0, 1)$, which proportionally reduces the transportation cost of flows via hub links. However, as noted by Campbell and O’Kelly [16], this classical assumption may not capture economies of scale correctly. O’Kelly and Bryan [42] recognized this weakness early on and proposed a more realistic approach by considering flow-dependent discounts for inter-hub links using a non-linear cost function. Later on, different approaches have been presented in the literature, including economies of scale with flow thresholds (Podnar et al. [45]), hub arc location models (Campbell et al. [14] and Campbell et al. [15]), and models with fixed and variable cost components (Kimms [29], O’Kelly et al. [41], and Serper and Alumur [48]). An extensive discussion on the importance and modeling of economies of scale is presented in Alumur et al. [4]. In this current paper, we apply both constant and flow-dependent discount factors for calculating the total transportation costs.

As hub location is a long-term strategic decision, incorporating uncertainty into hub location problems has lately received increased attention. There are several successful approaches addressing uncertainty in the literature including stochastic and robust optimization techniques. In stochastic optimization, the behavior of uncertain parameters are described by known probability distributions and the expected value of the objective function is evaluated based on these

distributions (e.g., Kleywegt et al. [30], Sim et al. [49], Contreras et al. [18], Mohammadi et al. [35], Adulyasak et al. [3], Sadeghi et al. [47] Correia et al. [19], Correia and Saldanha-da Gama [20], Peiró et al. [43], and Taherkhani et al. [52]). Robust optimization, on the other hand, works with a deterministic, set-based description of the uncertainty where there is no probabilistic information for the uncertain parameters. In such a case, uncertainty is described by a finite set of scenarios or is modeled assuming that the values of the uncertain parameters can change within predefined intervals (e.g., Meraklı and Yaman [33], Zetina et al. [55], de Sá et al. [22], and Ghaffarinasab [25]).) It is also possible to consider both approaches to model uncertainty (see e.g., Alumur et al. [7] and Taherkhani et al. [53]). Several studies that address uncertainty within different types of hub location problems have already shown that hub network designs can change significantly when uncertainty is incorporated into decision-making, see, for example, Contreras et al. [18], Alumur et al. [7], Zetina et al. [55], Peiro et al. [44], Taherkhani et al. [52], and Taherkhani et al. [53].

Another approach incorporating uncertainty in decision making is to integrate optimization and simulation techniques in which the simulation is used to evaluate the solution obtained from an optimization model. This methodology has been proposed for several problem types including production-distribution planning in supply chains (Lee et al. [32]), health care management (De Angelis et al. [21]), and distribution network design (Ko et al. [31]). In Ko et al. [31], they first propose a deterministic optimization formulation and use it to determine dynamic distribution network structures. They then apply the simulation model to capture the uncertainty in clients' demands, order-picking time, and travel time for the capacity plans of the warehouses based on service time and implement the approach to an example problem.

Acar et al. [2] propose a general hybrid solution methodology which can be adapted to various optimization problems, such as facility location, facility layout, and scheduling. The hybrid approach uses an initial value which is iteratively evaluated and adjusted by the simulation model. Following the idea proposed by Acar et al. [2], in this paper, a hybrid iterated optimization-simulation approach is developed to address uncertainty in hub location problems.

3 Deterministic Models

In this section, first, the problem setting is clearly defined and subsequently, the corresponding mathematical models for the single and multiple allocation cases are provided. Since uncertainty is not yet incorporated in those formulations, these will be referred to as the *deterministic* models.

3.1 Problem Setting

We model uncapacitated hub location problems with a cost minimization objective. The problem setting consists of a set of nodes and demand which has to be served between them. There is a set of potential hubs that may include all nodes. Both the number of hubs and the established connections between them are endogenously determined in the optimization process. Therefore, establishing a fully interconnected hub network is possible, but not required, and no particular inter-hub network structure is imposed a priori. Moreover, direct connections between two non-hub nodes are not allowed and all flow has to be routed via at least one hub facility. Regarding the design of the access network (collection and distribution connections), two separate models are developed to cover both the single and multiple allocation structures. In the latter case, no restrictions are set on how many hub facilities a demand node can be allocated to.

The objective of the presented hub location problem is the minimization of total costs while fully serving the existing demand of all origin-destination pairs. Because capacity restrictions are neither imposed on the hubs nor on the links, a feasible solution is always guaranteed, regardless of costs. Total costs consist of three separate parts: the variable transport costs for routing flow on the network, the fixed installation costs for opening hub facilities, and the fixed costs for establishing inter-hub links. The presented models minimize total costs by balancing the network design and transportation costs. There is economies of scale while using the inter-hub links due to bulk transportation on those links. In the basic formulation, it is assumed that the economies of scale is reflected on the transportation cost by a constant discount factor, which is independent of the amount of flow routed, on all inter-hub links. As it will be explained in Section 4.5, in the simulation phase, a flow-dependent discount factor is used for accurately calculating the total transportation costs.

The following notation is used for the parameters that are required for modeling the deterministic hub location problems:

N	Set of nodes.
ω_{ij}	Flow originated at node $i \in N$ which is destined to node $j \in N$.
$O_i = \sum_{j \in N} \omega_{ij}$	Total flow originated at node $i \in N$.
$D_i = \sum_{j \in N} \omega_{ji}$	Total flow destined to node $i \in N$.
f_k	Fixed set-up cost of locating a hub at node $k \in N$.
g_{kl}	Fixed cost of operating a hub link from hub $k \in N$ to hub $l \in N$.
c_{ij}	Cost of sending one unit of flow from node $i \in N$ to node $j \in N$.
α	Economies of scale discount factor for the flow transferred between hubs.

It is assumed that the values of all parameters are scaled considering the length of the planning horizon.

3.2 Single Allocation Model

The basic formulation for the single allocation hub location problem requires each non-hub node to be allocated to exactly one hub facility. A 3-index formulation based on the models introduced in Ernst and Krishnamoorthy [23] and Alumur et al. [5] is presented. Consequently, the decision variables are defined as follows:

$$\begin{aligned}
 x_{ik} &= \begin{cases} 1, & \text{if node } i \text{ is assigned to hub } k, \\ 0, & \text{otherwise. } (i, k \in N) \end{cases} \\
 &\quad (x_{kk} = 1 \text{ indicates that node } k \in N \text{ is a hub node}) \\
 z_{kl} &= \begin{cases} 1, & \text{if a hub link is operated from hub } k \text{ to hub } l, \\ 0, & \text{otherwise. } (k, l \in N) \end{cases} \\
 y_{ikl} &= \text{Amount of flow originated at node } i \text{ which is routed from hub } k \text{ to hub } l. \\
 &\quad (i, k, l \in N)
 \end{aligned}$$

Note that, in contrast to the model proposed by Alumur et al. [5], the decision variable for operating the inter-hub link z_{kl} depends on the chosen orientation. Therefore, the model can take different decisions regarding the operation of the hub arcs (k, l) and (l, k) .

The deterministic uncapacitated single allocation hub network design problem (SAHND) is modeled as:

$$\begin{aligned}
 \min \quad Z_{\text{SAHND}} &= \sum_{i \in N} \sum_{k \in N} c_{ik} O_i x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha c_{kl} y_{ikl} + \sum_{i \in N} \sum_{k \in N} c_{ki} D_i x_{ik} \\
 &\quad + \sum_{k \in N} f_k x_{kk} + \sum_{k \in N} \sum_{l \in N} g_{kl} z_{kl} \tag{1}
 \end{aligned}$$

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

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

$$\sum_{l \in N, l \neq k} y_{ikl} - \sum_{l \in N, l \neq k} y_{ilk} = O_i x_{ik} - \sum_{j \in N} \omega_{ij} x_{jk} \quad \forall i, k \in N \tag{4}$$

$$z_{kl} \leq x_{kk} \quad \forall k, l \in N \tag{5}$$

$$z_{kl} \leq x_{ll} \quad \forall k, l \in N \tag{6}$$

$$y_{ikl} \leq O_i z_{kl} \quad \forall i, k, l \in N \quad (7)$$

$$x_{ik} \in \{0, 1\} \quad \forall i, k \in N \quad (8)$$

$$z_{kl} \in \{0, 1\} \quad \forall k, l \in N \quad (9)$$

$$y_{ikl} \geq 0 \quad \forall i, k, l \in N \quad (10)$$

The objective function (1) accounts for the total costs. The first three terms calculate the transportation costs including collection, transfer, and distribution costs, respectively. The next two terms calculate the installation cost of hubs and the cost of operating hub links, respectively, which represent the network design costs.

Constraints (2) ensure that every node is allocated to exactly one hub facility. Constraints (3) guarantee that the demand nodes can be allocated only to open hubs. Constraints (4) represent the flow divergence (or flow balance) equations. For the flow originating at each node, the total flow entering into a hub node must be equal to the total outgoing flow. The left-hand-side of Constraints (4) calculates the flow going into and out of hub $k \in N$ on the inter-hub network, whereas the right-hand-side calculates the flow going into and out of hub k on the allocation connections both for the flow originating from node $i \in N$.

Constraints (5) and (6) ensure that an inter-hub link can only be established if hubs are opened at both end-nodes. Constraints (7) ensure that flow originating from node i can only be routed from hub k to hub l if an inter-hub link is established. Finally, constraints (8)-(10) define the domains of the decision variables.

3.3 Multiple Allocation Model

The formulation for the multiple allocation hub location problem allows non-hub nodes to be allocated to more than one hub facility. Although no upper limit is set, a minimum allocation to at least one hub is required. The presented model is based on the formulation proposed by Boland et al. [10] and expanded by incorporating inter-hub link costs proposed by Alumur et al. [5] for the single allocation version. The decision variables are defined as follows:

$$\begin{aligned}
\xi_k &= \begin{cases} 1, & \text{if node } k \text{ is a hub node,} \\ 0, & \text{otherwise. } (k \in N) \end{cases} \\
x_{ik} &= \text{Amount of flow originated at non-hub node } i \text{ sent directly to hub } k. \ (i, k \in N) \\
z_{kl} &= \begin{cases} 1, & \text{if a hub link is operated from hub } k \text{ to hub } l, \\ 0, & \text{otherwise. } (k, l \in N) \end{cases} \\
v_{ilj} &= \text{Amount of flow originated at node } i \text{ which is routed to node } j \text{ via hub } l \text{ as} \\
&\quad \text{the last hub on the route. } (i, l, j \in N) \\
y_{ikl} &= \text{Amount of flow originated at node } i \text{ which is routed from hub } k \text{ to hub } l. \\
&\quad (i, k, l \in N)
\end{aligned}$$

Note that similar to the single allocation case, in contrast to the model proposed by Alumur et al. [5], the decision variable for operating the inter-hub link z_{kl} depends on the chosen orientation. Therefore, the model can take different decisions regarding the operation of the hub arcs (k, l) and (l, k) . The deterministic uncapacitated multiple allocation hub network design problem (MAHND) is modeled as:

$$\begin{aligned}
\min \quad Z_{\text{MAHND}} &= \sum_{i \in N} \sum_{k \in N} c_{ik} x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha c_{kl} y_{ikl} + \sum_{i \in N} \sum_{l \in N} \sum_{j \in N} c_{lj} v_{ilj} \\
&\quad + \sum_{k \in N} f_k \xi_k + \sum_{k \in N} \sum_{l \in N} g_{kl} z_{kl} \tag{11}
\end{aligned}$$

$$s.t. \quad \sum_{k \in N} x_{ik} = O_i \quad \forall i \in N \tag{12}$$

$$\sum_{l \in N} v_{ilj} = \omega_{ij} \quad \forall i, j \in N \tag{13}$$

$$\sum_{l \in N, l \neq k} y_{ikl} - \sum_{l \in N, l \neq k} y_{ilk} = x_{ik} - \sum_{j \in N} v_{ikj} \quad \forall i, k \in N \tag{14}$$

$$x_{ik} \leq O_i \xi_k \quad \forall i, k \in N \tag{15}$$

$$x_{ik} \leq O_i (1 - \xi_i) \quad \forall i, k \in N, i \neq k \tag{16}$$

$$\sum_{i \in N} v_{ilj} \leq D_j \xi_l \quad \forall j, l \in N \tag{17}$$

$$v_{ilj} \leq D_j (1 - \xi_j) \quad \forall i, j, l \in N, l \neq j \tag{18}$$

$$z_{kl} \leq \xi_k \quad \forall k, l \in N \tag{19}$$

$$z_{kl} \leq \xi_l \quad \forall k, l \in N \tag{20}$$

$$y_{ikl} \leq O_i z_{kl} \quad \forall i, k, l \in N \tag{21}$$

$$\xi_k \in \{0, 1\} \quad \forall k \in N \quad (22)$$

$$z_{kl} \in \{0, 1\} \quad \forall k, l \in N \quad (23)$$

$$x_{ik} \geq 0 \quad \forall i, k \in N \quad (24)$$

$$y_{ikl} \geq 0 \quad \forall i, k, l \in N \quad (25)$$

$$v_{ilj} \geq 0 \quad \forall i, l, j \in N \quad (26)$$

The objective function (11) accounts for the transportation and network design costs structured similar to the single allocation version (1). As the flow originating at a node can be routed via different hubs in the multiple allocation case, different transportation costs on the allocation links have to be incorporated separately. Therefore, the aggregated term $O_i x_{ik}$ for the flow originating at node i is replaced by the actual flow on the collection links x_{ik} . The same holds true for the transport costs for distribution in which $D_i x_{ik}$ is replaced by v_{ilj} .

Constraints (12)-(14) represent the flow balance equations. Constraints (15) ensure that flow is only collected by hubs. This is extended by constraints (16) which state that every hub also collects its own flow. The same holds true for the distribution stated by constraints (17) and (18). Constraints (19) and (20) guarantee that an inter-hub link can only be established if a hub is opened at both incident locations. Due to constraints (21), flow originating from node i can only be routed from hub k to hub l if an corresponding inter-hub link is established between them. The remaining constraints (22)-(26) represent non-negativity and binary requirements.

4 Hybrid Solution Methodology

This section introduces a hybrid iterated optimization-simulation approach to address uncertainty in the above hub location models. In this approach, each deterministic problem (i.e., SAHND and MAHND) is first to be solved to optimality by using the expected values of the parameters. Next, the hub network solution obtained from the deterministic model is simulated under several scenarios by changing the values of the uncertain parameters to calculate an objective function value under uncertainty. The difference between the deterministic and simulated objective function value is called the impact of uncertainty on the objective. The impact of uncertainty is then to be incorporated into the proposed deterministic mathematical formulations, as will be explained later, to find a new candidate solution to be simulated. This process repeats until a previously simulated solution with uncertainty impact is found to be optimal in the deterministic formulation. The entire solution procedure is presented in Figure 1. As can be seen from this figure, the optimization model is solved to find a “good” candidate solution at each iteration, while the simulation is used as an evaluation method for the solution

obtained from this optimization model.

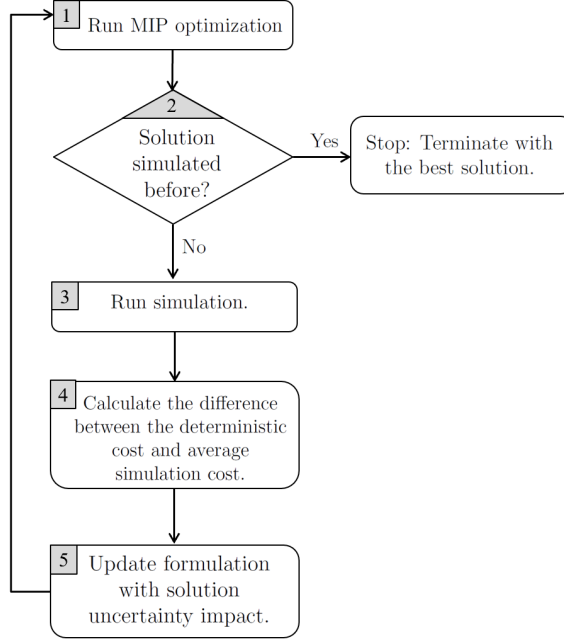


Figure 1: Generalized hybrid approach.

The first iteration consists of five steps: Initially, either one of the deterministic problems (i.e., SAHND or MAHND) is solved to optimality (step 1). At this point, uncertainty is not yet incorporated into the formulation. This obtained deterministic solution (i.e., the hub network), referred to as a scenario or candidate solution, is handed over to the simulation. Since no simulation run has been executed yet in the first iteration, the check whether a solution has been simulated before, consequently, always fails within the first iteration (step 2) and the process continues. The simulation tests the proposed optimal solution generated by the mathematical model by applying influence of uncertainty to the relevant parameters (e.g., demand and cost) (step 3).

After testing the candidate solution considering the uncertainty in the data, the difference between the objective values of optimization and simulation for the candidate solution is calculated. The resulting difference is the impact of uncertainty of the candidate solution (step 4). Finally, information on how a previously obtained candidate solution performed in the simulation is incorporated into the deterministic problem. This is achieved by adding a penalized cost term to the objective function as well as updating the relevant parameters in the mathematical formulation to be referred to as penalized formulations (step 5).

In the next iteration, the updated deterministic problem is solved again to optimality (step 1). This time, already simulated candidate solutions and their uncertainty impact are consid-

ered within the optimization phase. If a new candidate solution is found (step 2), the process continues the iteration with the simulation (step 3). This process iterates until the optimization (step 1) results in a candidate solution which has previously been found and simulated. In this case, the procedure stops and the currently considered solution is presented as the estimated optimal solution under uncertainty.

In order to function properly, the generalized iterative approach requires one major assumption: the total costs obtained under uncertainty in the simulation is assumed to be worse than the corresponding total costs obtained from the mathematical optimization. This assumption justifies the case where the optimal objective function obtained by deterministic data is disrupted once uncertainty is incorporated into the setting. Although a better or equal objective function value is also possible, we generate scenarios that results in an upper bound on the objective value to ensure that this assumption always holds true. This assumption may also justify a conservative decision maker attitude who wishes to be prepared for the worst-case scenarios when the problem setting is faced with uncertainty.

Following the hybrid iterated optimization-simulation approach, the impact of uncertainty is incorporated into the mathematical formulations presented in Section 3 by adding additional constraints and decision variables to the formulations. Moreover, an additional term, referred to as the penalty impact of already simulated candidate solutions, is added to the objective function of each model.

The required parameters that are valid for all penalized formulations are defined as follows:

r	Current iteration count-1.
M	A sufficiently large number (Big M).
$Z_{Simulation}^s$	Objective value of the simulation model of the candidate solution $s \in \{1..r\}$
$Z_{Optimization}^s$	Objective value of the optimization model of candidate solution $s \in \{1..r\}$.
P^s	Impact of uncertainty (penalty) of the candidate solution $s \in \{1..r\}$. ($P^s = Z_{Simulation}^s - Z_{Optimization}^s$)
Z_{min}	The minimum total simulation objective function obtained so far from any simulation run.

The objective value of the simulation model $Z_{Simulation}^s$ of candidate solution s is calculated by taking the expected value of the objective functions obtained under different scenarios.

Next, we present the penalized formulations for both single and multiple allocation versions where the impact of uncertainty is embedded in the models.

4.1 Single Allocation Model with Penalized Hub Network

This section presents penalized formulation for the single allocation version where the impact of uncertainty is incorporated into the formulation. Recall that due to the stopping criteria in the scenario-based hybrid approach explained in Section 4, the procedure repeats until the most recent simulated solution with uncertainty impact is found to be optimal in the current mathematical formulation.

The allocation structure of each obtained candidate solution can be represented by defining the following binary parameter:

$$h_{ik}^s = \begin{cases} 1, & \text{if node } i \text{ is assigned to hub } k \text{ in candidate solution } s, \\ 0, & \text{otherwise. } (i, k \in N, s \in \{1, \dots, r\}) \end{cases}$$

Note that this parameter also includes the information about the opened hubs as $x_{kk} = 1$ indicates that node $k \in N$ is a hub node (hub nodes are allocated to themselves). In order to correctly incorporate the impact of uncertainty P^s of the obtained candidate solution s in the model, an additional decision variable is required:

$$u^s = \begin{cases} 1, & \text{if candidate allocation structure } s \text{ has already been suggested in a previous} \\ & \text{iteration,} \\ 0, & \text{otherwise. } (s \in \{1, \dots, r\}) \end{cases}$$

When a previously obtained and simulated feasible solution s is considered again (i.e., $u^s = 1$), the presented decision variable ensures that the penalty obtained from the simulation run (impact of uncertainty) is added as a cost term to the objective function (Z) via $P^s u^s$. Therefore, the *uncapacitated single allocation hub network design problem with penalized allocation structure* (SAP) is modeled in the following form to find an unsimulated feasible solution with the best objective function value:

$$\begin{aligned} \min \quad Z_{\text{SAP}} = & \sum_{i \in N} \sum_{k \in N} c_{ik} O_i x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha c_{kl} y_{ikl} + \sum_{i \in N} \sum_{k \in N} c_{ki} D_i x_{ik} \\ & + \sum_{k \in N} f_k x_{kk} + \sum_{k \in N} \sum_{l \in N} g_{kl} z_{kl} + \sum_{s=1}^r P^s u^s \end{aligned} \quad (27)$$

$$s.t. \quad (2) - (10)$$

$$\sum_{i \in N} \sum_{k \in N} (2h_{ik}^s x_{ik} - x_{ik} - h_{ik}^s) \leq u^s - 1 \quad \forall s \in \{1, \dots, r\} \quad (28)$$

$$\sum_{i \in N} \sum_{k \in N} (2h_{ik}^s x_{ik} - x_{ik} - h_{ik}^s) \geq M(u^s - 1) \quad \forall s \in \{1, \dots, r\} \quad (29)$$

$$Z \leq Z_{min} \quad (30)$$

$$u^s \in \{0, 1\} \quad \forall s \in \{1, \dots, r\} \quad (31)$$

The former objective function of the basic single allocation model (1) is extended by the aforementioned cost term $P^s u^s$ for considering previously simulated candidate solutions and their impact of uncertainty in the model. While the constraints (2) - (10) of the basic model remain unchanged, additional ones are needed for the correct penalization of the previously simulated candidate solutions. Constraints (28) and (29) ensure the correct linkage between the basic decision variable for allocation x_{ik} and the parameter used to store the allocation structure of the candidate solutions h_{ik}^s . Only if these variables are completely identical ($x_{ik} = h_{ik}^s$ for all $i, k \in N$) for an existing and previously simulated candidate solution s , will the corresponding binary decision variable be set to $u^s = 1$ and, consequently, the penalty P^s be added in the objective function. Constraint (30) sets an upper bound for the objective function value Z . It is equal to the current best known objective value $Z_{Simulation}^s$ of any previously simulated candidate solution $s \in \{1, \dots, r\}$. This ensures that only new candidate solutions are considered in the optimization, which potentially are able to perform better in the simulation model than the current best solution.

4.2 Single Allocation Model with Penalized Hub Locations

Although the penalized single allocation model proposed in the previous section works well in the iterative scenario-based hybrid approach for solving small sized instances of the problem, one major disadvantage arises once the formulation is used on larger and more realistic problem settings. As candidate solutions are defined in the above model by their allocation structure, changing just one assignment between a non-hub node and a hub facility in the network already qualifies as a new solution. Combining this with the fact that the order in which new solutions are found depends on the initial deterministic data without uncertainty, the model will obtain many solutions with only minor adjustments in order to stay close to the first solution with minimal costs obtained without incorporating uncertainty. However, those changes only lead to small deviations in the simulation phase. Therefore, many iterations are needed before a major change occurs, such as changing hub locations or varying the total number of opened facilities, and a significant increase in the value of uncertainty can be observed.

In order to overcome this issue, an adapted formulation is proposed in which a candidate solution s is not defined by its allocation structure but only by its opened hubs. Hence, in the penalized version of the formulation, just the established hubs are reported in the candidate solution. In particular, for each candidate solution s , we fix the locations of the hubs obtained from the deterministic model and optimize for the rest of the decisions including the amount and routes of flows as well as the establishment of inter-hub links. Consequently, the binary parameter for storing the solutions is redefined as:

$$h_k^s = \begin{cases} 1, & \text{if node } k \text{ is a hub in candidate solution } s, \\ 0, & \text{otherwise. } (k \in N, s \in \{1, \dots, r\}) \end{cases}$$

The decision variable for incorporating the impact of uncertainty for candidate solution s is modified accordingly:

$$u^s = \begin{cases} 1, & \text{if candidate hub structure } s \text{ has already been suggested in a previous iteration,} \\ 0, & \text{otherwise. } (s \in \{1, \dots, r\}) \end{cases}$$

Therefore, constraints (28) and (29) in the SAP model are replaced with the following constraints:

$$\sum_{k \in N} (2h_k^s x_{kk} - x_{kk} - h_k^s) \leq u^s - 1 \quad \forall s \in \{1, \dots, r\} \quad (32)$$

$$\sum_{k \in N} (2h_k^s x_{kk} - x_{kk} - h_k^s) \geq M(u^s - 1) \quad \forall s \in \{1, \dots, r\} \quad (33)$$

No additional variables or parameters have to be added to the SAP formulation except for the above constraints (32) and (33), which ensure the correct linkage between the decision variable x_{kk} for opening hubs and the parameter used to store the hub network of the candidate solutions h_k^s . The objective function (27) and constraints (2)–(10) and (30)–(31) remain unchanged.

4.3 Multiple Allocation Model with Penalized Hub Locations

We define the candidate solutions in the multiple allocation model also by the hub locations rather than the hub network. The binary parameter for storing the solutions is defined as:

$$h_k^s = \begin{cases} 1, & \text{if node } k \text{ is a hub in candidate solution } s, \\ 0, & \text{otherwise. } (k \in N, s \in \{1, \dots, r\}) \end{cases}$$

The decision variable for correctly incorporating the impact of uncertainty P^s of the obtained

candidate solution s in the model is:

$$u^s = \begin{cases} 1, & \text{if candidate hub structure } s \text{ has already been suggested in a previous iteration,} \\ 0, & \text{otherwise. } (s \in \{1, \dots, r\}) \end{cases}$$

The *uncapacitated multiple allocation hub network design problem with penalized hub structure* (MAP) is then modeled as:

$$\begin{aligned} \min \quad Z_{\text{MAP}} = & \sum_{i \in N} \sum_{k \in N} c_{ik} x_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \alpha c_{kl} y_{ikl} + \sum_{i \in N} \sum_{l \in N} \sum_{j \in N} c_{lj} v_{ilj} \\ & + \sum_{k \in N} f_k \xi_k + \sum_{k \in N} \sum_{l \in N} g_{kl} z_{kl} + \sum_{s=1}^r P^s u^s \end{aligned} \quad (34)$$

$$s.t. \quad (12) - (26)$$

$$\sum_{k \in N} (2h_k^s \xi_k - \xi_k - h_k^s) \leq u^s - 1 \quad \forall s \in \{1, \dots, r\} \quad (35)$$

$$\sum_{k \in N} (2h_k^s \xi_k - \xi_k - h_k^s) \geq M(u^s - 1) \quad \forall s \in \{1, \dots, r\} \quad (36)$$

$$Z \leq Z_{\min} \quad (37)$$

$$u^s \in \{0, 1\} \quad \forall s \in \{1, \dots, r\} \quad (38)$$

Similar to the single allocation case, the former objective function of the basic multiple allocation model (11) is extended by the cost term $P^s u^s$ by taking the previously simulated candidate solutions and their impact of uncertainty into account leading to the new objective function (34). Constraints (12)-(26) of the basic multiple allocation model remain unchanged. Additional constraints (35) and (36) ensure the correct linkage between the basic decision variable for opening hubs ξ_k and the parameter used to store the hub network of the candidate solutions h_k^s . Constraint (37) sets the upper bound for the objective function value Z , while constraints (38) represent the binary requirement.

It is worth to mention that there is an alternative option to avoid visiting the same solution while solving a mixed-integer program using the commercial solver CPLEX. One may use the feature known as *Incumbent Callback* that can be used to simulate the cost of including uncertainty to produce a new upper bound, retain the best solution found so far, and then reject the incumbent integer solution to force CPLEX to keep searching. However, as will be explained in the next section, the proposed methodology in this study is more flexible in handling uncertainty

in different problem settings.

4.4 Simulation Models

In the simulation phase, the obtained candidate solutions are evaluated under the influence of uncertainty. Using a known distribution, a number of scenarios are generated for the uncertain parameters, and the objective value of the simulation model $Z_{Simulation}^s$ of candidate solution s is calculated by taking the expected value of the objective functions obtained under these different scenarios. In this step, for each candidate solution s , we fix the locations of the hubs obtained from the deterministic model and optimize for the rest of the decisions including the amount and routes of flows as well as the establishment of inter-hub links.

We include an additional parameter and constraint in the single allocation model SAHND to fix the value for opening hubs x_{kk}^s for the candidate solution s :

$$r_k^s = \begin{cases} 1, & \text{if node } k \text{ is a hub } (x_{kk}^s = 1) \text{ in the currently simulated candidate solution } s, \\ 0, & \text{otherwise. } (k \in N, s \in \{1, \dots, r\}) \end{cases}$$

$$x_{kk}^s = r_k^s \quad \forall k \in N, s \in \{1, \dots, r\}. \quad (39)$$

In the multiple allocation case, on the other hand, the hub location decisions are determined by the decision variable ξ_k . Hence, the opened hubs in the candidate solution s are fixed in the simulation model by changing ξ_k^s to a parameter:

$$\xi_k^s = \begin{cases} 1, & \text{if node } k \text{ is a hub in the currently simulated candidate solution } s, \\ 0, & \text{otherwise. } (k \in N, s \in \{1, \dots, r\}) \end{cases}$$

For each candidate solution s , both SAHND and MAHND are solved under different scenarios of uncertain parameters by fixing the locations of hubs, as described above, to obtain the simulated solution and its corresponding objective function value $Z_{Simulation}^s$. Then, the difference between the deterministic cost and expected value of the simulation cost is calculated and the methodology continues with the steps 4 and 5 as illustrated in Figure 1. The pseudocode of our generalized hybrid iterated optimization-simulation solution methodology is presented in Algorithm 1.

An alternative methodology to measure the impact of uncertainty on a solution is to combine the Taguchi loss function with simulation. In this technique, different scenarios using Monte Carlo simulation are generated and then, the expected value and the variance of the objective

Algorithm 1 Generalized hybrid approach.

```
1: for  $s \in \{1, \dots, r\}$  do
2:   SOLVE MIP optimization (i.e., SAHND or MAHND) and find  $Z_{Optimization}^s$ .
3:   if solution simulated before, then
4:     Break; terminate with the best solution.
5:   else
6:     Run the simulation (i.e., SAP or MAP) and find  $Z_{Simulation}^s$ .
7:     Calculate the difference  $P^s = Z_{Simulation}^s - Z_{Optimization}^s$ .
8:     Add the uncertainty impact ( $P^s$ ) to the objective function.
9:   end if
10: end for
```

function of the model is optimized under the generated scenarios (see, e.g., Abdolshah et al. [1] and Mohammadi et al. [34]). In the method proposed in this study, only the expected value of the objective function is calculated under the generated scenarios, but not the variance.

4.5 Flow-Dependent Economies of Scale

The classical hub location models use a fixed value α for discounting the inter-hub transportation. However, as discussed in Section 2, this approach may not correctly represent the economies of scale. The correct calculation of the discount from the economies of scale depends on the actual flows on the network connections. However, these actual flows can only be revealed after obtaining the optimal hub network.

In order to create a more realistic setting, in the simulation phase, flow-dependent economies of scale on both inter-hub and allocation links are included as proposed by Bryan [11]. Accordingly, the flow-dependent discount factor on each arc (k, l) is calculated as follows:

$$\alpha_{kl} = \begin{cases} 1 - \theta f_{kl}^\beta, & \text{if arc } (k, l) \text{ is an inter-hub link,} \\ 1 - 0.25 \theta f_{kl}^\beta, & \text{otherwise.} \end{cases} \quad (40)$$

where θ and β are referred to as the tuning coefficient and exponent, respectively, with $0 < \theta \leq 1$, $\beta > 0$. and f_{kl} is the resulting total flow on the arc (k, l) . In this calculation, the cost reductions on the allocation links are lower than that of the discounts on the inter-hub links for the same level of flows. This is due to the assumption proposed by Bryan [11] that the links for distribution and collection lack all the resources available to the inter-hub links and, as a consequence, cannot attain the same level of cost reduction.

To this end, for each candidate solution obtained from the deterministic model at iteration s , First, the amount of flow routed on each collection, distribution, and inter-hub link, referred to as

f_{kl} are calculated. For each inter-hub link (k, l) , $f_{kl} = \sum_{i \in N} y_{ilk}$ for both the single and multiple allocation models. For each collection link from node k to hub l , $f_{kl} = O_k$ for the single allocation models, whereas $f_{kl} = x_{kl}$ for the multiple allocation case. For each distribution connection from hub k to node l , on the other hand, $f_{kl} = D_l$ for single allocation and $f_{kl} = \sum_{i \in N} v_{ikl}$ for multiple allocation, respectively.

Then, in the simulation phase, the constant discount factor α incorporated into the transportation costs in the objective function is replaced with α_{kl} for the inter-hub links to account for the flow-dependent discount on arc (k, l) calculated by (40). For the allocation links, on the other hand, the discount factor α_{kl} is multiplied with the associated collection and distribution costs in the objective function.

We would like to note that the analysis that is performed with this flow-dependent discount function can be regarded as post-processing since the paths of flow do not change. Incorporating a more realistic calculation of economies of scale within the proposed models may require development of tailored solution methodologies, which can be an important avenue for future research.

5 Computational Results

In this section, extensive computational experiments is performed in order to analyze the performance of the presented scenario-based iterated optimization-simulation approach. For this purpose, the CAB dataset is used, which is customarily used in the literature and based on airline passenger interactions between 25 major cities in the United States (O’Kelly [39]). OR Library (Beasley [9]) provides the demand between each pair of cities (w_{ij}) and the corresponding transportation costs (c_{ij}). We use these values from the OR Library while solving all iterations with the deterministic models. As customarily done in the literature, the demand is scaled to sum up to one. Hub installation costs are calculated based on $f_k = 15 \log O_k$ where O_k represents the total flow originated at node $k \in N$ (Alumur et al. [7]). The demand flows from the CAB dataset result in values for the hub installation costs within the range $f_k \in [74.89, 92.41]$. Cost of operating an inter-hub link is set to 10% of the mean opening cost of the two incident hubs; i.e., $g_{kl} = 0.05(f_k + f_l)$ for all $k, l \in N$. The economies of scale discount factor is set to $\alpha \in \{0.4, 0.6, 0.8\}$, where $\alpha = 0.4$ is used as the default setting in the presented calculations unless stated otherwise.

All computational experiments were conducted on a computer with the following specifications: Intel Core i5-6500 3.20GHz CPU, 4 Cores and 16GB of RAM. The iterative approach was implemented in AnyLogic 7 University 7.3.2 and the integrated mathematical models were

solved by using IBM ILOG CPLEX 12.7.0.0. All instances were solved to optimality (10^{-5} gap) with increased precision on the feasibility tolerance (10^{-9}).

The main difficulty in solving the proposed formulations is that the size of the formulations grow extremely fast with the size of the instances. For example, the 25-node instances yield around 1,300 binary variables, 15,700 continuous variables, and 18,000 constraints for the single allocation model (1)-(10), and around 650 binary variables, 32,500 continuous variables, and 36,000 constraints for the multiple allocation model (11)-(26), both for their deterministic counterparts, respectively.

In what follows, we first present the results obtained by using deterministic data and then, report the solutions with uncertain demands. Next, the computational results for different sources of uncertainties are presented and analyzed. The final part of this section provides an analysis on the flow-dependent economies of scale discount factor.

5.1 Solutions Based on Deterministic Data

First, the models are solved with the deterministic data as a base for comparison. The computation times of solving instances to optimality for single and multiple allocation formulations are on average 15 and 18 minutes, respectively. We additionally solved the LP relaxations of the formulations and calculated the optimality gaps. These optimality gaps are on average around 6% and 4% for the single and multiple allocation formulations, respectively.

Figure 2 presents the resulting hub network structure for the single allocation case with different values of the economies of scale factor. The red squares depict the locations of the hub nodes and the red links show the inter-hub connections, whereas the blue links depict the allocation connections.

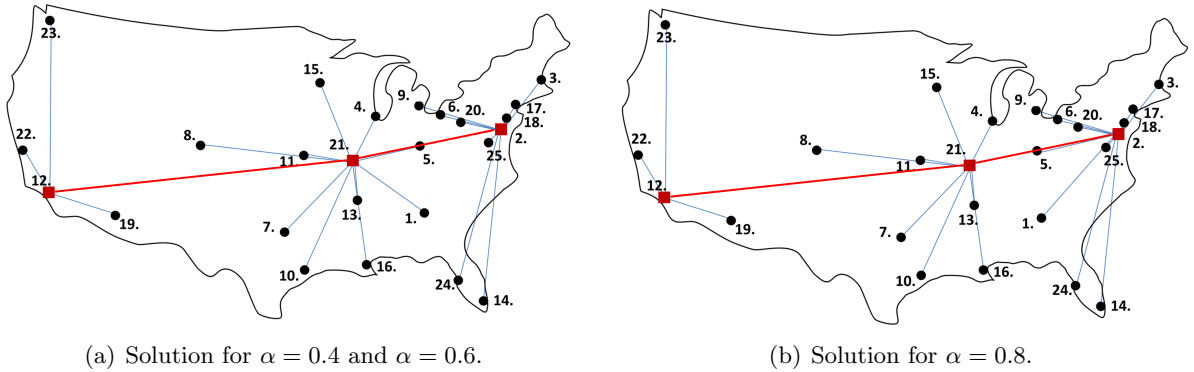


Figure 2: Single allocation solutions under deterministic data.

The solutions for the single allocation case with $\alpha = 0.4$ and $\alpha = 0.6$ turns out to be the same. Note that although the same hub network structure is established with $\alpha = 0.8$, the

allocations are different. The nodes 1 and 5 are allocated to the hub facility at node 2 with $\alpha = 0.8$, rather than to the hub at location 21.

The solutions for the multiple allocation case are depicted in Figure 3. The hub network structure, which consists of hubs located at nodes 12, 18, and 21, remains unchanged when the value for the economies of scale is increased from $\alpha = 0.4$ to 0.6. However, the number of allocation links increases. This effect is caused by the reduced discount on inter-hub links for an increase in α . When the inter-hub discount is reduced further to 0.8, this trend becomes even more apparent, which is depicted in Figure 3(c).

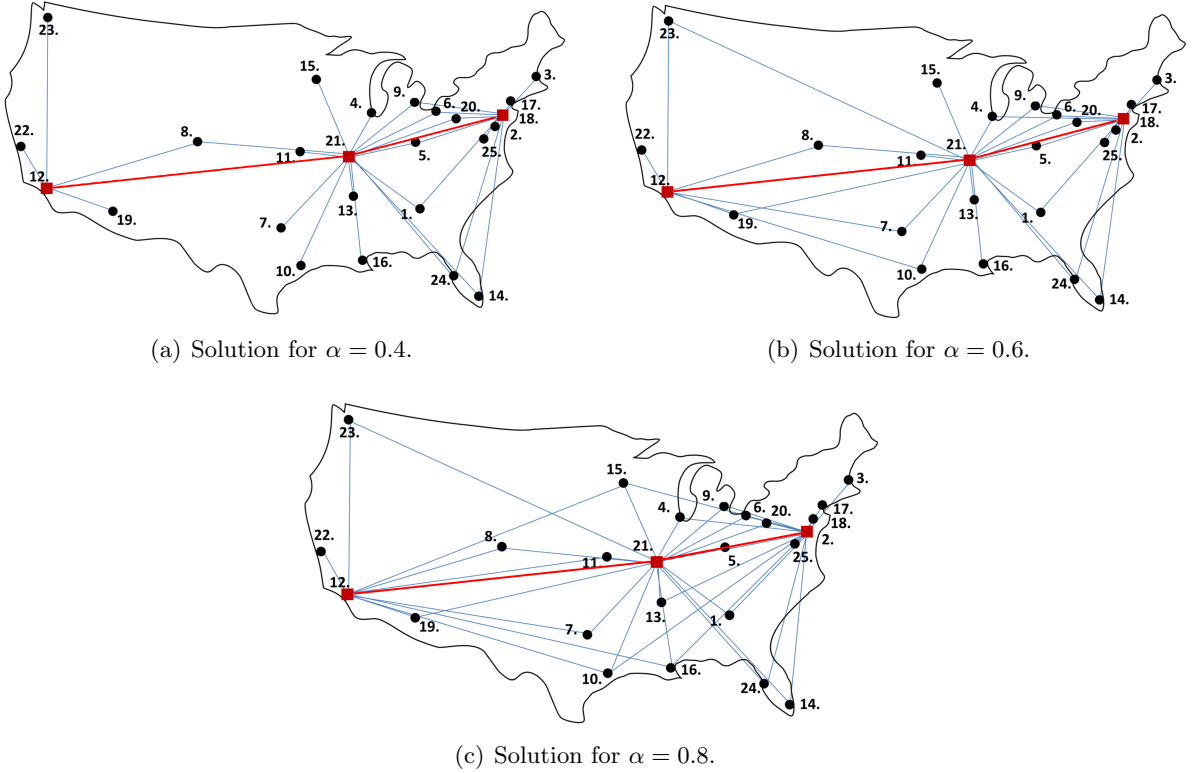


Figure 3: Multiple allocation solutions under deterministic data.

Comparing the solutions obtained under both the single and multiple allocation cases show that three major geographical regions of US are served by one hub each: the west coast, the mid-west, and the east coast. The value of the economies of scale parameter α has only a minor impact on the resulting hub network structure.

The resulting costs for the different solutions are presented in Table 1. The first two columns represent the allocation type and the economies of scale discount factor, respectively. The “Hub loc.” column presents the optimal hub locations obtained using the deterministic data. The columns “ $C_{\text{Transp.}}$ ”, “ $C_{\text{Hubs.}}$ ”, and “ $C_{\text{Links.}}$ ” correspond to the values of the total transportation costs, the set-up costs for opening the hub facilities, and the operational costs of the inter-hub

links, respectively. The last column reports the optimal total costs. Observe from the table that allowing multiple allocations for the same setting of α reduces the transportation costs. For both single and multiple allocation versions, an increase in α leads to a reduced discount and, therefore, to an increase in transportation costs.

Table 1: Costs in the deterministic solutions.

Allocation type	α	Hub loc.	$C_{\text{Transp.}}$	C_{Hubs}	C_{Links}	C_{Total}
Single	0.4	2,12,21	913.12	245.18	32.61	1190.91
	0.6	2,12,21	1041.92	245.18	32.61	1319.72
	0.8	2,12,21	1169.12	245.18	32.61	1446.91
Multiple	0.4	12,18,21	864.34	250.12	33.10	1147.56
	0.6	12,18,21	953.82	250.12	33.10	1237.04
	0.8	2,12,21	1026.84	245.18	32.61	1304.63

5.2 Uncertainty in Demands

The first set of the tests incorporates uncertainty only in the demands. For incorporating influence of uncertainty in demand, a uniform distribution on the interval $w_{ij} \in [0.01\bar{w}_{ij}, 10\bar{w}_{ij}]$ is used, where $\bar{w}_{ij}$ is the mean demand given in the CAB dataset (Alumur et al. [7]). It is important to note here that since both of the deterministic models refer to an uncapacitated setting, considering a stochastic demand is equivalent to solving its expected value problem, as shown in Contreras et al. [18] and Rostami et al. [46]. However, this is certainly not the case when the probability distribution of uncertain demand is not known apriori.

In the next section, we describe the stopping criteria for the iterated optimization-simulation approach.

5.2.1 Stopping criteria

Due to the fact that the hub location problem is NP-hard, solving large-scale instances can be time-consuming. Although the 25-node CAB network results in acceptable computation times in a deterministic setting, incorporating it in an iterative approach poses an additional challenge in computation. Each iteration consists of one optimization run of the deterministic hub location problem combined with additional runs of the penalized models in the simulation phase. As a consequence, performing consecutive iterations may easily lead to very high computation times. Depending on the parameter setting and the resulting upper bound, several hours are necessary until the iterative approach terminates. The convergence of the procedure might also be affected by the values of the parameters used in the optimization model. In the worst case, all variants enumeration might be required for reaching the stopping criteria. We acknowledge that

the iterated optimization-simulation is a heuristic approach and hence no convergence proof is possible. Using the 10-node subset from the CAB dataset, the long-term behavior of the iterated optimization-simulation approach is depicted in Figure 4.

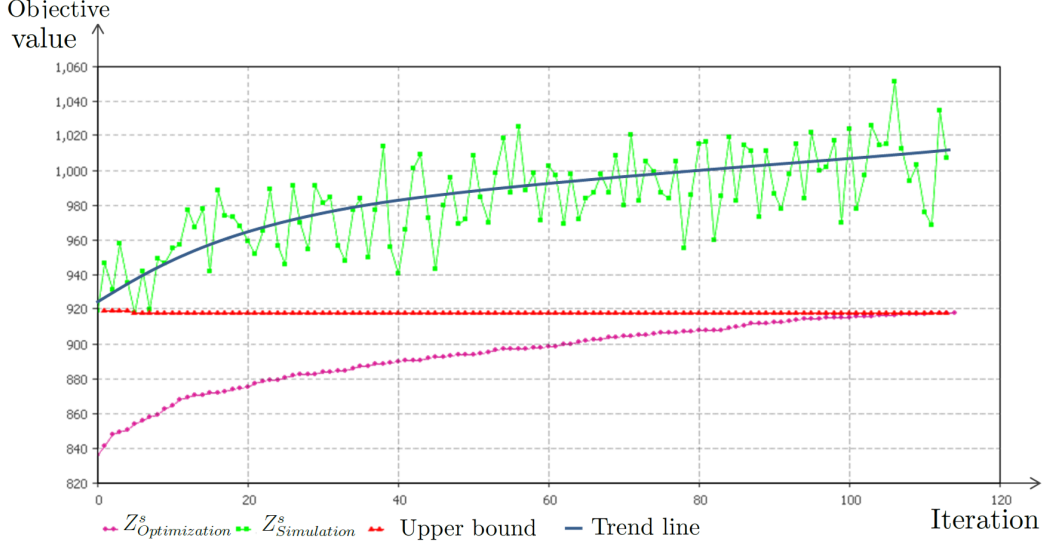


Figure 4: Long-term behavior of the iterated optimization-simulation approach.

The violet bottom line corresponds to the objective value of optimization $Z_{Optimization}^s$ for a candidate solution s . The green top line is the average objective value $Z_{Simulation}^s$ obtained by the candidate solution in the simulation phase. The red line represents the upper bound which is equal to the lowest obtained objective value in any simulation phase so far (i.e., Z_{min}). Even though the values for $Z_{Simulation}^s$ tend to fluctuate, the general structure follows an upward trend. This implies that the estimated optimal solution under uncertainty can be found in earlier iterations. Accordingly, the computation time for each run of the iterated optimization-simulation approach is limited to two hours. This also results in sufficient iterations for further analyses. Furthermore, since hub location is a long-term strategic decision, we believe that solving instances within two hours would be reasonable.

For the larger CAB dataset instances with 25 nodes, we compared the computation times spent on the simulation and optimization phases of the methodology. The computation time spent on the simulation phase is 5 minutes on average, while the rest is spent on the optimization phase. This implies that the computation time of the simulation phase is less than 5% of the total computation time. The results obtained for this set of instances show that using simulation can be promising and advantageous.

We also tested the performance of the algorithm using larger size instances with up to 40 nodes and 10 scenarios with and without a time limit. We observed that the behavior of

the algorithm does not change in larger size instances, i.e., the same trend as in Figure 4 is obtained. However, solving larger size instances is much more time consuming, particularly, for the optimization phase of the methodology.

5.2.2 Solutions with uncertain demands

We now present the solutions for the case where uncertainty is incorporated into the demand parameters. For the simulation phase, five scenarios for uncertain demands are generated and the expected value of the objective function $Z_{Simulation,average}^s$ is taken by weighting the results obtained in all five scenarios equally and, hence, taking the average, where s is the currently simulated candidate solution.

We separately solved the single and multiple allocation cases under each parameter setting and performed five different runs with a set of five different scenarios each for every $\alpha \in \{0.4, 0.6, 0.8\}$. Note that the scenarios used in the runs in the single allocation case differ from the ones in the multiple allocation case. In order to avoid biased results due to incidentally favorable scenarios, we increase variety in the simulation phase by generating new scenarios for every run.

To provide better insights on the proposed approach, the solutions obtained in the iterated optimization-simulation runs are compared with the solutions initially obtained when there is no uncertainty with the use of deterministic data (i.e., deterministic objective) as well as the results with their expected value counterparts (i.e., EVP). To obtain the solution for EVP, the corresponding deterministic models are solved using the average parameter values of the scenarios. A measure named *Value of Simulation* is defined to evaluate the benefits that can be obtained via the scenario-based hybrid approach and the significance of incorporating uncertainty in strategic decision making. The VoS-Det. is the relative difference between the objective functions obtained using only deterministic data but is now confronted with uncertainty ($Z_{Simulation}^{Initial}$) and the objective function obtained using iterated optimization-simulation technique ($Z_{Simulation}^*$), as defined in (41):

$$\text{VoS-Det.}(\%) = \frac{Z_{Simulation}^{Initial} - Z_{Simulation}^*}{Z_{Simulation}^*} \cdot 100 \quad (41)$$

Similarly, VoS-EVP is the relative difference between the objective functions obtained using the expected value problem (Z_{EVP}) and the objective function obtained using iterated optimization-simulation technique ($Z_{Simulation}^*$), as defined in (42):

$$\text{VoS-EVP}(\%) = \frac{Z_{EVP} - Z_{Simulation}^*}{Z_{Simulation}^*} \cdot 100 \quad (42)$$

In case either the deterministic model or the expected value problem results in the same objective value as in the final iteration of the scenario-based iterated optimization-simulation approach, no improvement has been made and, therefore, both the VoS-Det. and VoS-EVP will equal to 0%. Note that VoS-EVP might result in a negative value when a better objective function value is obtained with the expected value problem than iterated optimization-simulation.

The resulting values for the scenario-based iterated optimization-simulation approach with uncertain demands for both single and multiple allocation versions are listed in Table 2. The first two columns present the economies of scale factor and the number of the run, respectively. We refer to the solution with the deterministic data (i.e., base case) as the “Det.”. For each run in the single and multiple allocation cases, the “VoS” columns list the value of simulation compared with the deterministic and the EVP settings, respectively. The rest of the columns are as defined for Table 1. The rows “Avg. VoS” present the average value of simulation under five runs for each value of α in both the single and multiple allocation cases.

Table 2: Results with uncertain demands.

α	Single Allocation							Multiple Allocation							
	VoS[%]		Hub loc.	C _{Transp.}	C _{Hubs}	C _{Links}	C _{Total}	VoS[%]		Hub loc.	C _{Transp.}	C _{Hubs}	C _{Links}	C _{Total}	
	Det.	EVP						Det.	EVP						
0.4	Det.	-	-	2,12,21	913.1	245.2	32.6	1190.9	-	-	12,18,21	864.3	250.1	33.1	1147.6
	Run 1	0.5	0.15	4,12,18	902.5	258.2	34.7	1195.4	0.00	-0.08	12,18,21	864.3	250.1	33.1	1147.6
	Run 2	0.74	0.26	12,18,21	911.8	250.1	33.1	1195.8	0.00	0.11	12,18,21	864.3	250.1	33.1	1147.6
	Run 3	1.11	0.44	4,12,18	902.5	258.2	34.7	1195.4	1.35	0.14	4,12,18,24	773.5	336.2	50.8	1160.4
	Run 4	0.00	-0.10	2,12,21	913.1	245.2	32.6	1190.9	0.59	-0.05	4,12,18,24	773.5	336.2	50.8	1160.4
	Run 5	0.43	0.19	12,18,21	911.8	250.1	33.1	1195.1	0.29	0.17	4,12,18,24	773.5	336.2	50.8	1160.4
	Avg.	0.56	0.19						0.45	0.06					
0.6	Det.	-	-	2,12,21	1041.9	245.2	32.6	1319.7	-	-	12,18,21	953.8	250.1	33.1	1237.0
	Run 1	0.37	-0.22	12,18,21	1047.3	250.1	33.1	1330.5	0.00	0.04	12,18,21	953.8	250.1	33.1	1237.0
	Run 2	0.7	0.17	5,12,18	1048.6	246.1	32.3	1326.9	0.16	0.08	4,12,18	950.6	258.2	34.7	1243.5
	Run 3	0.58	0.26	2,4,12	1036.0	253.3	34.2	1323.5	0.22	-0.03	2,12,21	959.9	245.2	32.6	1237.7
	Run 4	0.35	-0.07	5,12,18	1048.6	246.1	32.3	1326.9	0.84	0.00	2,4,12	959.0	253.3	34.2	1246.5
	Run 5	0.15	0.08	2,4,12	1036.0	253.3	34.2	1323.5	0.51	0.07	4,12,18	950.6	258.2	34.7	1243.5
	Avg.	0.43	0.04						0.35	0.03					
0.8	Det.	-	-	2,12,21	1169.1	245.2	32.6	1446.9	-	-	2,12,21	1026.8	245.2	32.6	1304.6
	Run 1	0.15	0.09	2,5,12	1175.7	241.1	31.8	1448.6	0.33	0.26	12,18,21	1021.7	250.1	33.1	1304.9
	Run 2	0.38	-0.07	2,5,12	1175.7	241.1	31.8	1448.6	0.00	0.06	2,12,21	1026.8	245.2	32.6	1304.6
	Run 3	0.00	0.00	2,12,21	1169.1	245.2	32.6	1446.9	0.00	0.04	2,12,21	1026.8	245.2	32.6	1304.6
	Run 4	0.00	0.06	2,12,21	1169.1	245.2	32.6	1446.9	0.18	-0.12	12,18,21	1021.7	250.1	33.1	1304.9
	Run 5	0.41	0.28	2,5,12	1175.7	241.1	31.8	1448.6	0	0.00	2,12,21	1026.8	245.2	32.6	1304.6
	Avg.	0.19	0.07						0.10	0.05					

Observe from Table 2 that in 10 out of 15 runs, there is a positive value of simulation compared with the EVP, and in one instance they resulted in exactly the same solution ($\alpha = 0.8$, Run 3). The average VoS values for both deterministic and the EVP show that in these instances, the iterative scenario-based optimization-simulation approach results in solutions which perform

better under uncertainty than the initial solutions obtained under deterministic data as well as its expected value counterpart.

As shown in Table 2, the single allocation case outperforms the multiple allocation case in terms of VoS values when comparing the results for the same level of α . This is mainly because the established hub network and the corresponding allocation structure in the single allocation version using deterministic data (Figure 2) are inflexible regarding demand changes due to uncertainty. Every node is allocated to exactly one hub and the flow originating at this location has to be routed via the connected hub facility. Changes in the demand parameter do not allow for the use of additional allocation links. In the multiple allocation case, on the other hand, changes in demand can be partially compensated by making use of extra allocation connections and choosing the most cost-efficient routing option for each origin-destination pair. Because of this, uncertainty has a stronger impact on the single allocation solutions than on the multiple allocation ones. Hence, incorporating uncertainty via the iterated optimization-simulation approach leads to greater improvements in single allocation problems compared with the multiple allocation counterpart.

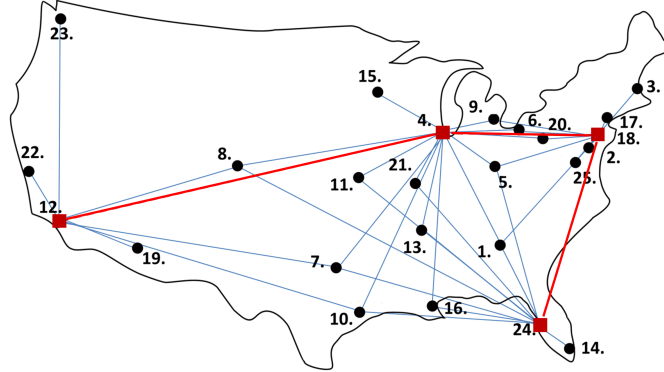


Figure 5: Hub network for the multiple allocation Run 3 in Table 2 with $\alpha = 0.4$.

Figure 5 depicts the hub network for multiple allocation Run3 in Table 2 with $\alpha = 0.4$ which results in the highest VoS value. In comparison to the base case solution depicted in Figure 3(a), which consists of three hubs and 30 allocation links, the solution proposed by the iterated optimization-simulation approach (Figure 5) extends the hub network to four hubs and establishes 38 allocation links. Additionally, the inter-hub network now consists of three instead of only two hub links. Using this structure, the solution obtained under uncertainty can use the extended hub network to partially outbalance demand fluctuations caused by uncertainty. Note that the solution with an additional hub at location 24 may not be optimal for all possible scenarios with demand under uncertainty, especially not for those that result in a reduction of the flow originating from nodes 14 and 24. However, using the average value $Z_{Simulation}^s$ based

on five randomly generated scenarios in the simulation phase, sufficiently prohibits the influence of randomness. This argument is also supported by the fact that the same hub locations are proposed frequently for the same parameter setting in different runs, each using a different set of five scenarios.

We next solved larger size instances with 40 nodes and 10 scenarios under demand uncertainty. The mean demand values (w_{ij}) and the transportation costs (c_{ij}) are from the extended CAB dataset (these values are readily available from the website: CAB100). The values of the rest of the parameters are calculated as described at the beginning of this section. For these set of instances, we used 10 scenarios, mainly because our preliminary results showed that using 10 scenarios is a reasonable compromise resulting in a good trade-off between computational time and incorporating the effects of uncertainty. The summary of the results for both single and multiple allocations is reported in Table 3. For each α value, the first row of this table presents the solution with the deterministic data and the second row is the result of our iterated optimization-simulation approach.

Table 3: Results of the instances with 40 nodes and 10 scenarios.

α	Single Allocation					Multiple Allocation				
	VoS[%]		Hub loc.	C_{Total}	Time (sec)	VoS[%]		Hub loc.	C_{Total}	Time (sec)
	Det.	EVP				Det.	EVP			
0.4	-	-	2,12,17,21,29,34,36	4,581.7	4,865	-	-	4,12,18,24,30	4,519.7	4,343
	3.74	1.67	2,12,18,21,29,30,36	4,759.6	58,824	2.84	1.19	12,18,21,24,29,36	4,651.8	56,520
0.6	-	-	4,12,18,21, 30,36	5,027.3	4,394	-	-	12,18,21,24,36	4,902.2	3,857
	2.63	0.81	2,12,18,24,30,36	5,163.1	51,300	2.31	-0.66	6,17,18,22,25,34	5,018.1	47,916
0.8	-	-	12,18,21,30,36	5,437.5	3,873	-	-	12,18,21,36	5,221.9	3,185
	1.08	-0.73	12,18,24,29,38	5,496.8	45,360	1.59	1.42	2,12,18,24,36	5,306.3	41,832

Similar to our observations from Table 2, the results in Table 3 underline the worth of the iterative optimization-simulation algorithm. In particular, the average VoS values for both single and multiple allocation versions show that the optimization-simulation approach results in solutions which perform better under uncertainty than the initial solutions obtained under deterministic data and its expected value counterpart. The average computational time of solving these set of instances with the algorithm is around 14 hours and the open number of hubs in the optimal solutions varies between five to seven. The computation times in Table 3 show that developing an efficient solution algorithm for this iterated optimization-simulation approach is necessary to be able to solve larger-size instances with more scenarios within a reasonable time frame.

5.3 Different Sources of Uncertainty

In this section, different sources of uncertainty are taken into account in the simulation phase. In addition to demand, uncertainty is also incorporated in costs including transportation cost (c_{ij}), hub installation cost (f_k), and operational cost for the inter-hub links (g_{kl}). For the transportation costs, values are generated from the interval $[0.95c_{ij}, 1.05c_{ij}]$. For the hub installation cost, values are generated from the interval $[10 \log O_k, 20 \log O_k]$ (Alumur et al. [7]). Accordingly, the costs for operating inter-hub links under uncertainty are derived from the interval $[0.035(f_k + f_l), 0.065(f_k + f_l)]$. To analyze the level of uncertainty in the solutions, left and right endpoints of the demand and cost intervals are multiplied with $(1 - \lambda)$ and $(1 + \lambda)$, respectively, as follows:

$$w_{ij} \in [0.01(1 - \lambda)\bar{w}_{ij}, 10(1 + \lambda)\bar{w}_{ij}]$$

$$c_{ij} \in [0.95(1 - \lambda)c_{ij}, 1.05(1 + \lambda)c_{ij}]$$

$$f_k \in [10(1 - \lambda) \log O_k, 20(1 + \lambda) \log O_k]$$

$$g_{kl} \in [0.035(1 - \lambda)(f_k + f_l), 0.065(1 + \lambda)(f_k + f_l)]$$

where $\lambda \in [0\%, 10\%, 20\%]$. Note that when λ increases, the level of uncertainty also increases.

The results for the scenario-based iterated optimization-simulation approach with uncertain demands and costs are presented in Table 4. The single and multiple allocation cases are solved separately under each parameter setting and performed five different runs with a set of five different scenarios each for every level of uncertainty $\lambda \in \{0\%, 10\%, 20\%\}$. $\alpha = 0.4$ is set for all runs. The solutions obtained in the iterated optimization-simulation runs are compared with the solutions obtained when uncertainty is not included and only deterministic data is used (i.e., Det.) as well as their expected value counterparts (i.e., EVP).

The VoS percentages obtained with both deterministic and the EVP are reported in Table 4. The results show that the solutions obtained by the iterative scenario-based iterated optimization-simulation approach yield a positive value of simulation in more than 80% of the test runs, implying that the obtained hub networks perform better under uncertainty compared with the corresponding solutions obtained with deterministic and average data. A higher level of uncertainty $\lambda = 20\%$ results in a higher value of simulation. This applies to both single and multiple allocation versions. On the other hand, when we compare the single and multiple allocation results, the single allocation version on average outperforms the multiple allocation in terms of VoS.

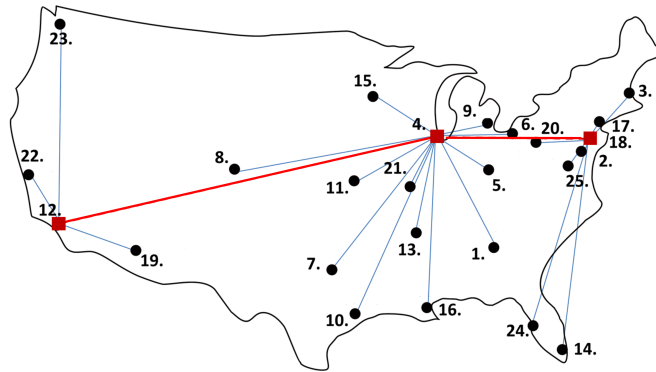
The results provided in Tables 2 and 4 show that all of the average values of simulation in Table 4 are higher than that of the ones presented in Table 2 when $\alpha = 0.4$. Since the results

Table 4: Results with different sources of uncertainty.

λ		Single Allocation							Multiple Allocation						
		VoS[%]		Hub loc.	$C_{Transp.}$	C_{Hubs}	C_{Links}	C_{Total}	VoS[%]		Hub loc.	$C_{Transp.}$	C_{Hubs}	C_{Links}	C_{Total}
		Det.	EVP						Det.	EVP					
0%	Det.	-	-	2,12,21	913.1	245.2	32.6	1190.9	-	-	12,18,21	864.3	250.1	33.1	1147.6
	Run 1	1.57	0.89	4,12,18	902.5	258.2	34.7	1195.4	1.19	-0.35	2,12,21	874.9	245.2	32.6	1152.7
	Run 2	1.37	0.63	5,12,17	912.6	256.2	33.3	1202.1	0.79	0.58	2,12,21	874.9	245.2	32.6	1152.7
	Run 3	3.77	1.97	5,12,18	921.7	246.1	32.3	1200.0	2.8	0.64	4,12,18	862.1	258.2	34.7	1155.0
	Run 4	1.92	-0.74	2,4,12	909.9	253.3	34.2	1197.4	1.5	0.74	4,12,18	862.1	258.2	34.7	1155.0
	Run 5	2.07	0.95	2,4,12	909.9	253.3	34.2	1197.4	1.37	-0.72	2,12,21	874.9	245.2	32.6	1152.7
	Avg.	2.14	0.74						1.53	0.18					
10%	Run 1	2.42	0.81	4,12,17	903.8	268.3	35.7	1207.9	2.8	1.36	2,12,21	874.9	245.2	32.6	1152.7
	Run 2	2.71	0.65	5,12,17	912.6	256.2	33.3	1202.1	3.49	2.00	4,12,18	862.1	258.2	34.7	1155.0
	Run 3	2.38	0.33	5,12,17	912.6	256.2	33.3	1202.1	2.55	1.08	4,12,18	862.1	258.2	34.7	1155.0
	Run 4	1.86	-0.26	2,4,12	909.9	253.3	34.2	1197.4	1.44	1.53	2,12,21	874.9	245.2	32.6	1152.7
	Run 5	2.76	1.66	4,12,17	903.8	268.3	35.7	1207.9	0.66	-1.19	2,12,21	874.9	245.2	32.6	1152.7
	Avg.	2.43	0.64						2.19	0.96					
	Run 1	1.3	0.90	5,12,18	921.7	246.1	32.3	1200.0	5.34	2.32	4,12,18	862.1	258.2	34.7	1155.0
20%	Run 2	5.01	2.40	4,12,18	902.5	258.2	34.7	1195.4	2.82	-0.66	4,12,18	862.1	258.2	34.7	1155.0
	Run 3	6.56	1.98	4,12,18	902.5	258.2	34.7	1195.4	1.33	0.85	2,12,21	874.9	245.2	32.6	1152.7
	Run 4	6.25	2.08	5,12,18	921.7	246.1	32.3	1200.0	2.27	1.92	2,12,21	874.9	245.2	32.6	1152.7
	Run 5	2.57	-0.56	2,4,12	909.9	253.3	34.2	1197.4	2.69	0.74	2,12,21	874.9	245.2	32.6	1152.7
	Avg.	4.34	1.36						2.89	1.03					

for the iterated optimization-simulation approach provided in Table 4 contain more uncertain parameters, the benefit of the hybrid approach for providing more reliable solutions is clearly more significant.

Figure 6 presents the hub network for the single allocation run 13 from Table 4, where $\lambda = 20\%$ and $\alpha = 0.4$. When we compare the resulting hub network with the base case (Figure 2(a)), we observe that locations of the two hubs has changed. The hub located at node 2 has moved to 18 and the hub at node 21 has moved to node 4. Moreover, the allocations of two of the nodes has been switched from the east coast hub to the central hub. Consequently, allocating more nodes to the central hub, which serves as a waypoint between the two peripheral hubs, results in a higher value of simulation once uncertainty is incorporated into the model (i.e., VoS=6.56%).

Figure 6: Hub network for the single allocation Run 13 in Table 4 with $\alpha = 0.4$.

The results from the analysis on the iterative scenario-based optimization-simulation approach, particularly the VoS-Det. values show that ignoring uncertainty in the demands and costs leads to sub-optimal solutions in all test runs. The associated increase in costs, represented by the value of simulation, may be up to 6.56% in the single, and 5.34% in the multiple allocation cases. However, we would like to acknowledge that all results are instance based and may depend on the dataset used for computational experiments.

5.4 Solutions with Flow-Dependent Economies of Scale

All the computational results presented in the previous section were performed using a fixed value (α) for discounting the inter-hub transportation. In this section, we analyze the impact of using a flow-dependent economies of scale discount factor in the simulation models. The discounts for collection, distribution, and inter-hub links are calculated through function (40) and these values are then incorporated into the simulation models. We set $\theta = 0.5$ and $\beta = 0.8$ in order to satisfy the condition $(\beta + 1) \leq \frac{1}{\theta}$ proposed by Horner and O’Kelly [28].

The resulting values for the scenario-based iterated optimization-simulation approach with various sources of uncertainty and flow-dependent economies of scale are listed in Table 5. We solved the single and multiple allocation cases separately for low level of uncertainty $\lambda = 0\%$ and performed five different runs with a set of five different scenarios each. For the deterministic solution, $\alpha = 0.4$ is taken. $\Delta C_{\text{Total}} [\%]$ columns present the percentage of cost increase between iterated optimization-simulation runs and the deterministic solution.

Table 5: Results with flow-dependent economies of scale and different sources of uncertainty.

	Single Allocation								Multiple Allocation							
	VoS [%]		Hub loc.	$C_{\text{Transp.}}$	C_{Hubs}	C_{Links}	C_{Total}	$\Delta C_{\text{Total}} [\%]$	VoS [%]		Hub loc.	$C_{\text{Transp.}}$	C_{Hubs}	C_{Links}	C_{Total}	$\Delta C_{\text{Total}} [\%]$
	Det.	EVP							Det.	EVP						
Det.	-	-	2,12,21	913.1	245.2	32.6	1190.9	-	-	-	12,18,21	864.3	250.1	33.1	1147.6	-
Run 1	1.02	0.32	12,18,21	911.8	250.1	33.1	1195.1	0.35	1.26	0.47	2,12,21	874.9	245.2	32.6	1152.7	0.45
Run 2	4.04	1.85	2,4,12	909.9	253.3	34.2	1197.4	0.55	1.83	0.87	2,12,21	874.9	245.2	32.6	1152.7	0.45
Run 3	2.77	0.90	5,12,17	912.6	256.2	33.3	1202.1	0.94	2.94	1.11	4,12,18	862.1	258.2	34.7	1155.0	0.65
Run 4	3.67	1.45	2,4,12	909.9	253.3	34.2	1197.4	0.55	2.08	0.94	2,12,21	874.9	245.2	32.6	1152.7	0.45
Run 5	2.46	0.99	5,12,17	912.6	256.2	33.3	1202.1	0.94	3.21	1.84	4,12,18	862.1	258.2	34.7	1155.0	0.65
Avg.	2.79	1.10							2.26	1.05						

The results in Table 5 show that a positive value of simulation (VoS) for both deterministic and EVP is obtained in every test run. Furthermore, the average values of simulation, which are 2.79% and 1.1% for the single and 2.26% and 1.05% for the multiple allocation case, indicate that the approach incorporating flow-dependent economies of scale results in solutions superior to the ones with constant discount factors provided in Table 4 for an equivalent value of uncertainty level, i.e., $\lambda = 0\%$.

Figure 7 shows the hub network for single allocation Run 2 from Table 5 with flow-dependent economies of scale. Comparing the network with the base case (Figure 2(a)), we observe that

the hub located at node 21 in the base case is now replaced by node 4 (Figure 7). All allocation connections turned out to be the same with the base case except for the nodes 1 and 9.

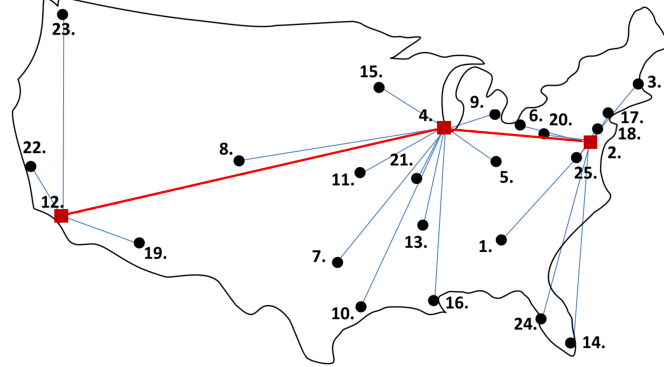


Figure 7: Hub network for the single allocation Run 2 in Table 5 with flow-dependent economies of scale.

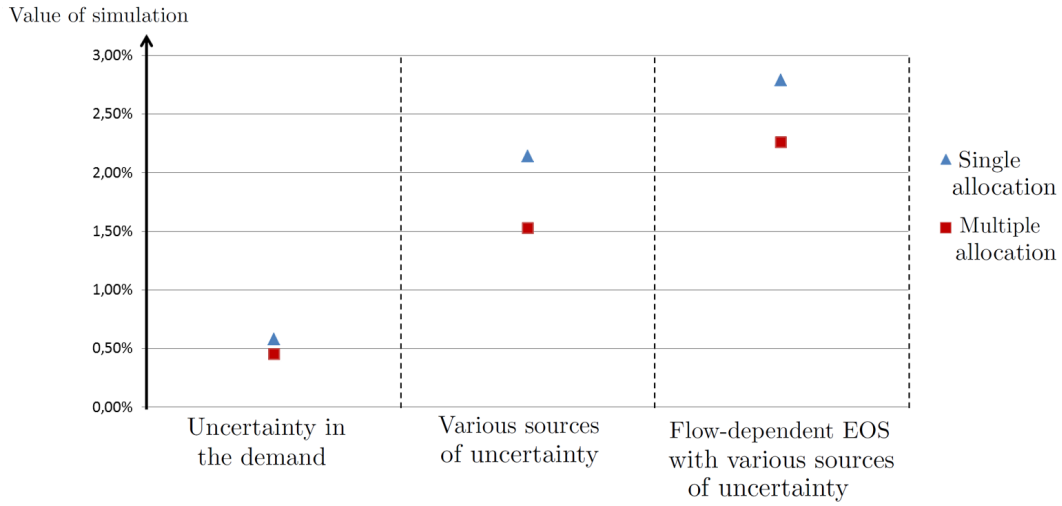


Figure 8: Value of simulation for different types of uncertainty.

In order to further demonstrate the value of simulation generated by the iterative scenario-based optimization-simulation approach, the percent changes in costs obtained under uncertainty are also stated in Table 5. Observe from Table 5 that the percentage of cost increase is up to 0.94% in the single and 0.65% in the multiple allocation cases. One may see this increase in costs in the iterated optimization-simulation approach as a drawback. However, comparing the percentages of cost increase ($\Delta C_{\text{Total}}[\%]$) with the value of simulation ($\text{VoS}[\%]$), which in return represents the decrease in costs under uncertainty compared to the base case solution, the improvement by using the proposed approach becomes evident. For example, observe from Table 5 that the maximum value for $\Delta C_{\text{Total}}[\%]$ is 0.94% that resulted from the runs 3 and

5 in the single allocation case. However, these increases are compensated by an average cost reduction of 2.77% and 2.46% over all randomly generated scenarios under uncertainty for runs 3 and 5, respectively. This compensation increases up to 4.04% in run 2 for the single allocation case.

Figure 8 gives an aggregated overview of the average value of simulation obtained for all types of uncertainty for setting $\alpha = 0.4$ ($\theta = 0.5$, $\beta = 0.8$ for the flow-dependent case) and $\lambda = 0\%$. In accordance with the trends observed in Tables 2, 4 and 5, incorporating additional types of uncertainty and flow-dependent economies of scale leads to an increased value of simulation. Consequently, we expect the approach to perform even better if additional sources of uncertainty are incorporated.

6 Conclusion

In this paper, we studied single and multiple allocation versions of the uncapacitated hub network design problem under various types and levels of uncertainty. Based on the generalized hybrid solution methodology, we developed an iterative scenario-based optimization-simulation approach, which obtains good quality solutions for both versions of the problem. In order to measure the performance of the methodology, we introduced the value of simulation, which compares the obtained solutions using deterministic data or the expected value problem, with the solutions obtained under uncertainty.

We conducted extensive computational analysis with more than 70 test runs and 350 scenarios on the CAB data set. For this purpose, we generated scenarios that include stochastic influence in the demands and are extended to consider uncertainty in transportation costs as well as in fixed costs for both establishing the hub facilities and operating the inter-hub links. Additionally, we allowed not only for varying sources but also for different levels of uncertainty in the generation of these scenarios. In order to correctly incorporate economies of scale, we performed a set of runs with flow-dependent discounts in the simulation phase.

The results show that the optimal hub network initially obtained under deterministic data changes significantly when uncertainty is considered. Using the value of simulation for evaluating the performed computational tests, we observed that incorporating an increased number of sources of uncertainty in the model results in higher values of simulation. The iterative scenario-based simulation optimization approach always outperformed the deterministic optimal solution when more than one type of uncertainty is considered. The highest performance is achieved when flow-dependent economies of scales is included in the approach. We also observed that the value of simulation increases when the level of uncertainty goes up.

To find the boundaries of our methodology, we solved larger-sized instances with a higher number of nodes and scenarios. We were able to solve instances with up to 40-nodes and 10 scenarios. However, the computational times were quite large, which were around 14 hours. Accordingly, the development of an efficient solution methodology (either a heuristic or an exact one) for the proposed iterated optimization-simulation approach to solve practical and larger-size instances is definitely worth pursuing in the future.

This paper is the first attempt to integrate optimization with simulation for hub network design and the results demonstrate the potential value of using this methodology for decision-making under uncertainty. Even though we implemented this methodology for the classical formulations, the mathematical models can readily be extended to more complex settings, such as multiple modes of transportation, multiple time periods, and the inclusion of direct links. This paper can also be considered a “proof-of-concept” for further and more complex joint implementations of optimization and simulation. Future work may incorporate other realistic features within the simulation phase, such as non-linear transportation cost functions, reliability of hubs, congestion, and user behavior, to design more resilient hub networks.

References

- [1] Abdolshah, M., Yusuff, R. M., Ismail, M. Y. B., and Hong, T. S. (2009). Process capability analysis using Monte Carlo simulation. In *2009 International Conference on Information Management and Engineering*, pages 335–339. IEEE.
- [2] Acar, Y., Kadipasaoglu, S., and Day, J. (2009). Incorporating uncertainty in optimal decision making: Integrating mixed integer programming and simulation to solve combinatorial problems. *Computers & Industrial Engineering*, 56(1):106–112.
- [3] Adulyasak, Y., Cordeau, J.-F., and Jans, R. (2015). Benders decomposition for production routing under demand uncertainty. *Operations Research*, 63(4):851–867.
- [4] Alumur, S. A., Campbell, J. F., Contreras, I., Kara, B. Y., Marianov, V., and O’Kelly, M. E. (2021). Perspectives on modeling hub location problems. *European Journal of Operational Research*, 291:1–17.
- [5] Alumur, S. A., Kara, B. Y., and Karasan, O. E. (2009). The design of single allocation incomplete hub networks. *Transportation Research Part B: Methodological*, 43(10):936–951.
- [6] Alumur, S. A., Kara, B. Y., and Karasan, O. E. (2012a). Multimodal hub location and hub network design. *Omega*, 40(6):927–939.

- [7] Alumur, S. A., Nickel, S., and Saldanha-da Gama, F. (2012b). Hub location under uncertainty. *Transportation Research Part B: Methodological*, 46(4):529–543.
- [8] Bakker, H., Dunke, F., and Nickel, S. (2020). A structuring review on multi-stage optimization under uncertainty: Aligning concepts from theory and practice. *Omega*, 96:102080.
- [9] Beasley, J. (1990). OR Library: Distributing test problems by electronic mail. <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/phubinfo.html>. Accessed on October 24, 2017.
- [10] Boland, N., Krishnamoorthy, M., Ernst, A. T., and Ebery, J. (2004). Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research*, 155(3):638–653.
- [11] Bryan, D. (1998). Extensions to the hub location problem: Formulations and numerical examples. *Geographical Analysis*, 30(4):315–330.
- [12] Campbell, J. F. (1992). Location and allocation for distribution systems with transshipments and transportation economies of scale. *Annals of Operations Research*, 40(1):77–99.
- [13] Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72(2):387–405.
- [14] Campbell, J. F., Ernst, A. T., and Krishnamoorthy, M. (2005a). Hub arc location problems: Part I - Introduction and Results. *Management Science*, 51(10):1540–1555.
- [15] Campbell, J. F., Ernst, A. T., and Krishnamoorthy, M. (2005b). Hub arc location problems: Part II - Formulations and Optimal Algorithms. *Management Science*, 51(10):1556–1571.
- [16] Campbell, J. F. and O’Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Science*, 46(2):153–169.
- [17] Carson, Y. and Maria, A. (1997). Simulation optimization: methods and applications. In *Proceedings of the 29th conference on Winter simulation*, pages 118–126.
- [18] Contreras, I., Gordeau, J., and Laporte, G. (2011). Stochastic uncapacitated hub location. *European Journal of Operational Research*, 212:518–528.
- [19] Correia, I., Nickel, S., and Saldanha-da Gama, F. (2018). A stochastic multi-period capacitated multiple allocation hub location problem: Formulation and inequalities. *Omega*, 74:122–134.
- [20] Correia, I. and Saldanha-da Gama, F. (2019). Facility location under uncertainty. In *Location science*, pages 185–213. Springer.

- [21] De Angelis, A., Felici, G., and Impelluso, P. (2003). Integrating simulation and optimisation in health care centre management. *European Journal of Operational Research*, 150:101–114.
- [22] de Sá, E. M., Morabito, R., and de Camargo, R. S. (2018). Benders decomposition applied to a robust multiple allocation incomplete hub location problem. *Computers & Operations Research*, 89:31–50.
- [23] Ernst, A. T. and Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p -hub median problem. *Location Science*, 4(3):139–154.
- [24] Ernst, A. T. and Krishnamoorthy, M. (1998). Exact and heuristic algorithms for the uncapacitated multiple allocation p -hub median problem. *European Journal of Operational Research*, 104(1):100–112.
- [25] Ghaffarinasab, N. (2018). An efficient matheuristic for the robust multiple allocation p -hub median problem under polyhedral demand uncertainty. *Computers & Operations Research*, 97:31–47.
- [26] Grigoryev, I. (2015). *AnyLogic 7 in three days: A quick course in simulation modeling*. AnyLogic North America, Charleston, SC.
- [27] Heckmann, I. (2016). *Towards Supply Chain Risk Analytics: Fundamentals, Simulation, Optimization*. Springer Gabler, Wiesbaden.
- [28] Horner, M. and O’Kelly, M. (2001). Embedding economies of scale concepts for hub network design. *Journal of Transport Geography*, 9(4):255–265.
- [29] Kimms, A. (2006). Economies of scale in hub & spoke network design models: We have it all wrong. *Perspectives on operations research*, pages 293–317.
- [30] Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502.
- [31] Ko, H. J., Ko, C. S., and Kim, T. (2006). A hybrid optimization/simulation approach for a distribution network design of 3pls. *Computers & Industrial Engineering*, 50(4):440–449.
- [32] Lee, Y. H., Kim, S. H., and Moon, C. (2002). Production-distribution planning in supply chain using a hybrid approach. *Production Planning & Control*, 13(1):35–46.
- [33] Meraklı, M. and Yaman, H. (2016). Robust intermodal hub location under polyhedral demand uncertainty. *Transportation Research Part B: Methodological*, 86:66–85.

- [34] Mohammadi, M., Siadat, A., Dantan, J.-Y., and Tavakkoli-Moghaddam, R. (2015). Mathematical modelling of a robust inspection process plan: Taguchi and Monte Carlo methods. *International Journal of Production Research*, 53(7):2202–2224.
- [35] Mohammadi, M., Torabi, S., and Tavakkoli-Moghaddam, R. (2014). Sustainable hub location under mixed uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 62:89–115.
- [36] Nguyen, A.-T., Reiter, S., and Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. *Applied energy*, 113:1043–1058.
- [37] Nickel, S., Schöbel, A., and Sonneborn, T. (2001). Hub location problems in urban traffic networks. In *Mathematical Methods on Optimization in Transportation Systems*, pages 95–107. Springer.
- [38] O’Kelly, M. E. (1986). The location of interacting hub facilities. *Transportation Science*, 20(2):92–106.
- [39] O’Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, 32(3):393–404.
- [40] O’Kelly, M. E., Bryan, D., Skorin-Kapov, D., and Skorin-Kapov, J. (1996). Hub network design with single and multiple allocation: A computational study. *Location Science*, 4(3):125–138.
- [41] O’Kelly, M. E., Campbell, J. F., de Camargo, R. S., and de Miranda Jr, G. (2015). Multiple allocation hub location model with fixed arc costs. *Geographical Analysis*, 47(1):73–96.
- [42] O’Kelly, M. E. and Bryan, D. (1998). Hub location with flow economies of scale. *Transportation Research Part B: Methodological*, 32(8):605–616.
- [43] Peiró, J., Corberán, Á., Martí, R., and Saldanha-da Gama, F. (2019). Heuristic solutions for a class of stochastic uncapacitated p -hub median problems. *Transportation Science*, 53(4):1126–1149.
- [44] Peiro, J., Corberan, A., Marti, R., and Saldanha-da Gama, F. (2019). Heuristic solutions for a class of stochastic uncapacitated p -hub median problems. *Transportation Science*, page <https://www.researchgate.net/publication/326010181>.
- [45] Podnar, H., Skorin-Kapov, J., and Skorin-Kapov, D. (2002). Network cost minimization using threshold-based discounting. *European Journal of Operational Research*, 137(2):371–386.

- [46] Rostami, B., Kämmerling, N., Naoum-Sawaya, J., Buchheim, C., and Clausen, U. (2021). Stochastic single-allocation hub location. *European Journal of Operational Research*, 289(3):1087–1106.
- [47] Sadeghi, M., Jolai, F., Tavakkoli-Moghaddam, R., and Rahimi, Y. (2015). A new stochastic approach for a reliable p-hub covering location problem. *Computers & Industrial Engineering*, 90:371–380.
- [48] Serper, E. Z. and Alumur, S. A. (2016). The design of capacitated intermodal hub networks with different vehicle types. *Transportation Research Part B: Methodological*, 86:51–65.
- [49] Sim, T., Lowe, T. J., and Thomas, B. W. (2009). The stochastic p -hub center problem with service-level constraints. *Computers & Operations Research*, 36(12):3166–3177.
- [50] Skorin-Kapov, D., Skorin-Kapov, J., and O’Kelly, M. (1996). Tight linear programming relaxations of uncapacitated p -hub median problems. *European Journal of Operational Research*, 94(3):582–593.
- [51] Taherkhani, G. and Alumur, S. A. (2019). Profit maximizing hub location problems. *Omega*, 86:1–15.
- [52] Taherkhani, G., Alumur, S. A., and Hosseini, M. (2020). Benders decomposition for the profit maximizing capacitated hub location problem with multiple demand classes. *Transportation Science*, 54(6):1446–1470.
- [53] Taherkhani, G., Alumur, S. A., and Hosseini, M. (2021). Robust stochastic models for profit-maximizing hub location problems. *Transportation Science*.
- [54] Yaman, H. (2009). The hierarchical hub median problem with single assignment. *Transportation Research Part B: Methodological*, 43(6):643–658.
- [55] Zetina, C. A., Contreras, I., Cordeau, J.-F., and Nikbakhsh, E. (2017). Robust uncapacitated hub location. *Transportation Research Part B: Methodological*, 106:393–410.