Computers & Operations Research 38 (2011) 559–570

Contents lists available at ScienceDirect

Computers & Operations Research

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

Single-allocation ordered median hub location problems ☆

J. Puerto [a], A.B. Ramos [a], A.M. Rodríguez-Chía [b,*]

[a] Facultad de Matemáticas, Universidad de Sevilla, Spain
[b] Facultad de Ciencias, Universidad de Cádiz, Spain

ARTICLE INFO

Available online 29 July 2010

Keywords:
Hub location problems
Ordered median function

ABSTRACT

The discrete ordered median location model is a powerful tool in modeling classic and alternative location problems that have been applied with success to a large variety of discrete location problems. Nevertheless, although hub location models have been analyzed from the sum, maximum and coverage point of views, as far as we know, they have never been considered under an alternative unifying point of view. In this paper we consider new formulations, based on the ordered median objective function, for hub location problems with new distribution patterns induced by the different users' roles within the supply chain network. This approach introduces some penalty factors associated with the position of an allocation cost with respect to the sorted sequence of these costs. First we present basic formulations for this problem, and then develop stronger formulations by exploiting properties of the model. The performance of all these formulations is compared by means of a computational analysis.

1. Introduction

The literature of hub location covers a large variety of models where the main goal is to minimize some globalizing function of the operation costs. See the surveys [1,9] and the references therein. Despite that most papers have been devoted to the minimization of the overall transportation cost (sum) (see [4,7,8,11,13,14,17,24–29,37,40], among others), in some cases other objectives have also been taken into account. Among them we mention the minimization of the largest transportation cost and the coverage cost [5,10,20,21,23,31,38,39]. However, these models have never been considered under alternative points of view as required by nowadays logistics. Practitioners need, more and more, flexible models that incorporate the different roles of the parties in the supply chain network. Roughly speaking, the classical hub location models try to minimize the sum of the transportation costs of each origin–destination path, i.e. the system is analyzed from the logistics provider point of view, see [41]. However, depending on the driving force of the logistics network we can distinguish alternative points of views, as for instance, suppliers, clients or a combination of both points of views. A first attempt to deal with this flexibility has been already addressed with some success in certain classical location models (see e.g. [2,15,16,18,19,41]) and it is mainly concerned with the differentiation of users' roles within the supply chain network or equivalently with the use of new or different distribution patterns (origin–destination delivery paths) in the distribution phase. In this paper, we elaborate on the direction mentioned above and present a model that allows to differentiate on the role played by the different parties in a hub-type supply chain network. First, we incorporate flexibility through rank dependent compensation factors. Second, we assume that the driving force in the supply chain is shared by the suppliers and the distribution system; suppliers support the transportation costs from the origin sites to the first hub and the distribution system supports the transportation cost from the first hub to the destination sites.

In this paper, we introduce and analyze hub location problems with a new type of distribution pattern induced by the different roles of the users within the supply chain. Actually, we consider supply chain networks where the operations within the network consist of segmented origin–destination deliveries of known amounts of a commodity. The goal is to simultaneously make decisions on the location of the intermediate transhipment points (hubs) and on the origin–destination delivery paths (distribution patterns). We seek to establish an intermediate distribution system with a fixed number of hubs that minimizes the overall operation cost of the supply chain network. Moreover, we assume that any origin–destination delivery path is composed of, at most, two components: (1) the subpath that goes from an origin site to the first access point (first hub) to the distribution system, and (2) the subpath that links first hubs to final destinations. In addition, this last component is itself divided into two parts: (2.1) the inter-hubs link and (2.2) the link from the last hub to the final destination. This structure allows us to distinguish between different cost generating entities according to the roles played

☆ This research was partially supported by "Ministerio de Ciencia e Innovación" (research projects MTM2007-67433-C02-01,02, MTM2010-19576-C02-01,02, DE2009-0057), "Fundación Séneca" (research project 08716/PI/08), and "Junta de Andalucía" (research projects P06-FQM-01364, P06-FQM-01366), Spain.
* Corresponding author. Tel.: +34 956 016087; fax: +34 956 016050.
E-mail address: antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

0305-0548/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2010.07.018

within the supply chain. On the one hand, we assume that each origin must support the cost to reach the first hub (i.e. the distribution system), while the intermediate distribution system supports the remaining delivery costs, namely the cost induced once the commodity has reached the first hub (from the first hub to the final destination), i.e., the distribution system is the responsible for any commodity once it reaches the first hub. It is somewhat similar to parcel companies where urban/first level franchises carry the commodities to the assigned distribution centers of the parcel company and then, the company delivers the product to its final destination.

Each one of the components of any origin–destination delivery path described above gives rise to a cost that is weighted by different compensation factors depending on the role of the party that supports the cost. From an application point of view, consolidated deliveries within the distribution system, i.e. between hubs, and from the last hub to the final destinations, should be cheaper than the first component of the cost since they can be done using larger vehicles or cheaper transportation modes (due for instance to the larger size of the distribution system which implies a dominant position in price negotiation). Therefore, in our models we assume that the inter-hub links are covered by the same type of vehicles (planes, big trucks, etc.) and the costs associated with these links have a fixed discount $0<\alpha<1$. Moreover, we also assume that the links between last-hub and the destination sites are also covered by a different type of vehicle (small trucks, vans, etc.) and the associated costs have another discount factor $0<\delta<1$. A possible reason of using two different discount factors for the inter-hub links and the links between the last-hub and destination sites may be due to the different loading and unloading systems used in the hubs with respect to the ones used in the destination sites (the location of hubs implies the implementation of automatic loading/unloading systems that allow to handle big trucks, while destination sites may use traditional methods that only works with small vehicles).

In addition, deliveries from the origin sites to the distribution system are scaled by rank dependent weights. We assume that the commodity of each origin is transported to a single (unique) first hub that represents the access point to the distribution system. These weights can be seen as compensation factors that try to diminish unfair situations of the origin sites with respect to the distribution system. The reader may note that we are simultaneously making decisions on placing hubs that define the intermediate distribution system, and establishing the delivery paths from origin sites to final destination. Thus, a solution that is good for the system (the entire supply chain) might not be acceptable for single parties if in that solution their costs to reach the system are too high relative to similar costs for other parties. In this situation some compensation to unhappy sites may be expected to prevent those sites from not using the system. The goal of our rank dependent weights is to compensate unfair situations, such as those described above. For instance, if a solution places a set of hubs so that the accessibility cost of origin j is greater than the corresponding cost of origin j', the model tries to favor j with respect to j' assigning weights $\lambda_j \le \lambda_{j'}$. (Note that these weights do not penalize site j' but instead they compensate site j because these lambdas reduce the dispersion of the costs.) In order to incorporate this ordinal information in the overall transportation cost, the objective function applies a correction factor to the transportation cost of the commodity that is sent from each origin to a first hub (to reach the system) which is dependent on the position of that cost relative to similar costs from other origin sites. For example, a different penalty might be applied if the transportation cost of the commodity from origin site j was the 5th-most expensive cost rather than the 2nd-most expensive, see [3,30,32,33,35,36]. It is even possible to neglect some origin by assigning a zero penalty. This adds a "sorting"-problem to the underlying hub location problem, making formulation and solution much more challenging.

Our goal is to present a unified framework to analyze these and related models of hub location that consider ordinal information to cope with actual requirements from logistics. Formally, the objective is to minimize the total transportation cost of the flows between each origin–destination pair, routed through at most two hubs, once we have applied rank dependent compensation factors on the transportation costs of the origin-first hub links, and fixed scaling factors for the inter-hub and hub-final destination transportation costs.

The rest of the paper is organized as follows: In Section 2 we describe formally the model and provide a mathematical programming formulation using variables with three indices. Section 3 studies alternative formulations for this hub location problem using covering variables. In Section 4, we present a preliminary computational analysis to determine the limits of solving the problem with the 3-index variable and covering variable formulations using standard MIP solvers. Section 5 develops some improvements and strengthening with respect to the previous formulations and an alternative formulation under the hypothesis of λ-weights given in non-decreasing order. These improvements are computationally compared in Section 6. The paper ends with some conclusions. In the Appendix, we provide a result stating that some constraints presented above, that in the general case are required to get a valid formulation, are redundant whenever the cost structure satisfies the triangular inequality.

2. The model and the 3-index formulation

Let A denote a given set of N client sites and identify these with integers $1,\ldots,N$. Each site is collecting or gathering some commodity that must be sent to the remaining sites. Let $w_{jm} \ge 0$ be the amount of commodity to be supplied from the j th- to the m th-site for all $j,m \in \{1,\ldots,N\}$ and let $W_j = \sum_{m=1}^{N} w_{jm}$. In the following, we assume without loss of generality that the set of candidate sites for establishing hubs is identical to the set of sites A. Let $c_{jm} \ge 0$ denote the unit cost of sending commodity from site j to site m (not necessarily satisfying the triangular inequality). We assume that $c_{jj}=0$, $\forall j=1,\ldots,N$. Let $p \le N$ be the number of hubs to be located and $X \subset A$ with $|X|=p$ denote a feasible set of candidate sites. A solution for the problem is a feasible set of candidate sites X, plus a set of paths connecting pairs (flow patterns) of sites j, m for all $j,m \in \{1,\ldots,N\}$ in such a way that each path traverses at least one and no more than two hubs from X. To be more precise, (i) if the origin site j and the destination site m are not hubs, the flow must go through one or two intermediate hubs; (ii) if either the origin or the destination sites are hubs, the flow between them can be either directly sent or sent through an additional hub; and (iii) if both origin and destination sites are hubs, the flow must go directly from the origin to the destination.

In addition, this model compensates origin site-first hub transportation costs by using parameters $\lambda=(\lambda_1,\ldots,\lambda_N)$. These scaling factors will be assigned to the origins depending on the order of the sequence of transportation costs of the commodity with the same origin to the first hub. Indeed, if a solution sends the commodity from the origin site j via a first hub k and this delivery cost, namely $W_j c_{jk}$, was ranked in the ith position among these type of costs then this term would be scaled by λ_i, i.e. the corresponding objective function component would be $\lambda_i W_j c_{jk}$. In addition, we also consider a compensation parameter $0<\alpha<1$ for the deliveries between hubs and another parameter $0<\delta<1$, $\alpha<\delta$ for the deliveries between hubs and final destination sites. These parameters may imply that, at times, using a second hub results in a cheaper connection than going directly from the first hub to the final destination.

Observe that depending on the choices of the λ-vector we can obtain different criteria to account for the costs from the origins to their first hubs in the objective function. For instance, if $\lambda=(0,\ldots,0,1,\overset{k}{\ldots},1)$, were considered the objective function would be the sum of the k-largest costs (k-centrum). In the following, we illustrate the different behavior of our model for different choices of the λ-vector and the single allocation p-hub median. For this purpose we use the CAB data set publicly available at http://people.brunel.ac.uk/~mastjjb/jeb/info.html (see [34]), with $N=20$, $p=5$, $\alpha=0.7$ and $\delta=1.2\alpha$. Indeed, Table 1 reports the locations of the hub sites for different criteria (p-center, i.e., $\lambda=(0,\ldots,0,1)$, trimmed mean with $k_1=k_2=8$, i.e., $\lambda=(0,\overset{8}{\ldots},0,1,1,1,1,0,\overset{8}{\ldots},0)$, k-centrum with $k=9$, i.e., $\lambda=(0,\ldots,0,1,\overset{9}{\ldots},1)$) and the locations of the hubs sites for the single allocation p-hub median. We can check that the solutions we obtained are different for all the considered criteria. Moreover, we observe, from Table 1, that even in those cases where the same site is chosen as a hub with respect to different criteria, the allocation pattern of origin sites to hubs is different. For instance, in the optimal solution for the center and the single allocation p-hub median there are two hubs at sites 2 and 13. However, the allocations are different. Indeed, $14\longrightarrow 2$ for the center problem and $14\longrightarrow 13$ for the single allocation p-hub median model.

2.1. 3-Index formulation

A natural way to attack the formulation of the above model may be using variables that keep tracks of the order of the transportation costs from each origin to its first hub and the entire path followed by the flow between each origin–destination pair. This approach would give rise to a formulation with 5-index variables, one for the order and the remaining four indices, for the origin–destination paths. Indeed, based on the fact that any origin–destination path cannot traverse more than two intermediate hubs we can use four indices to define the flow patterns. The first and fourth indexes stand for the origin and destination sites whereas the second and third ones indicate the intermediate hubs. Rather than this formulation, our first model decouples the requirement of the sorting from the pattern followed by each flow and then, a formulation with *only* 3-index variables can be stated. In order to formulate this model we consider a set of λ-weights, where λ_i can be seen as a correction factor to the i th-position with $i=1,\ldots,N$. In addition, we define the following set of variables:

$$r^i_{jk}=\begin{cases}1 & \text{if the flow from the origin site } j \text{ goes first to the hub } k \text{ and } W_j c_{jk} \text{ is the } i\text{th lowest value of the transportation costs from each origin to its first hub}\\ 0 & \text{otherwise}\end{cases}$$

$x_{k\ell m}=$ flow that goes through a first hub k and a second hub ℓ with destination m

$$y_k=\begin{cases}1 & \text{if a hub is located at the site } k\\ 0 & \text{otherwise}\end{cases}$$

with $i,j,k,\ell,m=1,\ldots,N$.

Thus, the 3-index formulation is

$$\min \sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N \lambda_i c_{jk} r^i_{jk} W_j+\sum_{k=1}^N\sum_{\ell=1}^N\sum_{m=1}^N x_{k\ell m}(\alpha c_{k\ell}+\delta c_{\ell m}) \tag{1}$$

$$\text{s.t.} \quad \sum_{i=1}^N\sum_{k=1}^N r^i_{jk}=1, \quad \forall j=1,\ldots,N \tag{2}$$

$$\sum_{j=1}^N\sum_{k=1}^N r^i_{jk}\le 1, \quad \forall i=1,\ldots,N \tag{3}$$

$$\sum_{i=1}^N\sum_{j=1}^N r^i_{jk}\le N y_k, \quad \forall k=1,\ldots,N \tag{4}$$

$$\sum_{i=1}^N r^i_{jj}=y_j, \quad \forall j=1,\ldots,N \tag{5}$$

$$\sum_{\ell=1}^N x_{k\ell m}=\sum_{i=1}^N\sum_{j=1}^N r^i_{jk} w_{jm}, \quad \forall k,m=1,\ldots,N \tag{6}$$

$$\sum_{j=1}^N\sum_{k=1}^N r^i_{jk} c_{jk} W_j \le \sum_{j=1}^N\sum_{k=1}^N r^{i+1}_{jk} c_{jk} W_j, \quad \forall i=1,\ldots,N-1 \tag{7}$$

$$x_{k\ell m}\le (1-y_m)\sum_{j=1}^N w_{jm}, \quad \forall k,\ell,m=1,\ldots,N,\ \ell\neq m \tag{8}$$

$$\sum_{\ell=1}^N\sum_{m=1}^N x_{k\ell m}\le y_k\sum_{j=1}^N W_j, \quad \forall k=1,\ldots,N \tag{9}$$

$$\sum_{k=1}^N\sum_{m=1}^N x_{k\ell m}\le y_\ell\sum_{j=1}^N W_j, \quad \forall \ell=1,\ldots,N \tag{10}$$

$$\sum_{k=1}^N y_k=p \tag{11}$$

$$r^i_{jk}\in\{0,1\}, \quad x_{k\ell m},\ y_k\ge 0, \ \forall i,j,k,\ell,m=1,\ldots,N \tag{12}$$

The objective function accounts for the weighted sum of the three components of the shipping cost, namely origin site to first hub, inter-hub connections and last hub to final destination site. The first block of shipping costs is accounted after the compensation process using the lambda parameters, whereas the second and third blocks are scaled with the α and δ parameters, respectively. Constraints (2) ensure that each origin site j is allocated exactly to one position in the ordered sequence of transportation costs and all the flow from the origin site j is also associated with a unique first hub. Constraints (3) guarantee that any position in the sorted vector of *origin site-first hub* costs is

Table 1
Hub locations and allocation pattern for the data instance of the CAB problem $N=20$ and $p=5$.

Center	Trimmed mean	k-Centrum	p-Hub median
$2,3,14\longrightarrow 2$	$1,10,13,16\longrightarrow 1$	$1,14,16\longrightarrow 1$	$2,3,17,18\longrightarrow 2$
$4,11,15,19\longrightarrow 4$	$4,7,8,11,12,15,17\longrightarrow 4$	$4,5,6,9,11,13,15\longrightarrow 4$	$4,5,6,9,20\longrightarrow 5$
$8,12\longrightarrow 8$	$3,14,18\longrightarrow 18$	$7,8,10\longrightarrow 7$	$8,11,15\longrightarrow 11$
$7,10,13,16\longrightarrow 13$	$19\longrightarrow 19$	$12,19\longrightarrow 12$	$1,7,10,13,14,16\longrightarrow 13$
$1,5,6,9,17,18,20\longrightarrow 20$	$2,5,6,9,20\longrightarrow 20$	$2,3,17,18,20\longrightarrow 17$	$12,19\longrightarrow 19$

allocated to at most one. Constraints (4) ensure that one origin may be allocated to a specific first hub only if it is open. Observe that this family of constraints can be presented in a disaggregated form, i.e.

$$\sum_{i=1}^{N} r_{jk}^{i} \le y_k, \quad \forall j,k=1,\ldots,N$$

However, our computational experience shows that using the disaggregated form provides worse computational running times than using the original constraints (3). Therefore, we have kept them in the aggregated form. Constraints (5) state that if an origin j is a hub itself, then the first hub in any of its origin–destination paths must be j.

Constraints (6) are flow conservation constraints and they ensure that the flow that enters any hub k with final destination m is the same that the flow that leaves hub k with destination m. Constraints (7) give us the order relationship of the shipping costs between the origin sites-first hubs. Constraints (8) ensure that if the final destination site is a hub, then the flow goes at most through one additional hub. Constraints (9) and (10) establish that the intermediate nodes in any origin–destination path should be open hubs. Finally, constraint (11) fixes the number of hubs to be located.

By constraints (5), variables y_k can be defined as continuous since r-variables will force them to take $0-1$ values. In fact, variables y can be removed of this formulation using constraints (5), but we have kept them in order to clearly distinguish the location variables.

Note that the family of constraints (5) and (8) are redundant whenever the cost structure satisfies the triangular inequality; however, they are useful in reducing solution times (see the Appendix for further details).

3. An alternative formulation based on covering variables: covering 3-index formulation

In this section we provide an alternative formulation to the previous one based on covering variables (see [6,12,15,22,30]). In order to do that, we first define G as the number of different non-zero elements of the cost sequence $W_j c_{jk}$ for any $j,k=1,\ldots,N$. Hence, we can order the different values of this sequence in non-decreasing sequence:

$$c_{(0)} := 0 < c_{(1)} < c_{(2)} < \cdots < c_{(G)} := \max_{1 \le j,k \le N} \{W_j c_{jk}\}$$

Given a feasible solution, we can use this ordering to perform the sorting process of the allocation costs. This can be done by the following covering variables ($i=1,\ldots,N$ and $h=1,\ldots,G$):

$$u_{ih} := \begin{cases} 1 & \text{if the } i\text{th smallest allocation cost is at least } c_{(h)} \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$

The ith smallest allocation cost is equal to $c_{(h)}$ if and only if $u_{ih}=1$ and $u_{i,h+1}=0$.

Clearly, the first part of the objective function, as appears for instance in (1), can be equivalently written as

$$\sum_{i=1}^{N} \sum_{h=1}^{G} \lambda_i \cdot (c_{(h)} - c_{(h-1)}) \cdot u_{ih}$$

In order to formulate this model, we use the following set of variables:

$$r_{jk} = \begin{cases} 1 & \text{if the commodity sent from origin} \\ & \text{site } j \text{ goes first to the hub } k \\ 0 & \text{otherwise} \end{cases}$$

Observe that the relationship between these variables and the r_{jk}^{i}-variables defined in the 3-index formulation is $r_{jk} = \sum_{i=1}^{N} r_{jk}^{i}$.

Therefore the constraints that establish the link between variables u and r are

$$\sum_{i=1}^{N} u_{ih} = \sum_{j=1}^{N} \sum_{\substack{k=1 \\ W_j c_{jk} \ge c_{(h)}}}^{N} r_{jk}, \quad \forall h=1,\ldots,G$$

These constraints state that the number of allocations with a cost at least $c_{(h)}$ must be equal to the number of sites that support shipping costs to the first hub greater than or equal to $c_{(h)}$. Moreover, we impose the following group of sorting constraints on the u_{ih}-variables:

$$u_{ih} \le u_{i+1,h} \quad i=1,\ldots,N-1, \; h=1,\ldots,G$$

Applying the covering variables to the 3-index formulation, we obtain the *covering 3-index formulation*:

$$\min \quad \sum_{i=1}^{N} \sum_{h=2}^{G} \lambda_i (c_{(h)} - c_{(h-1)}) u_{ih} + \sum_{k=1}^{N} \sum_{\ell=1}^{N} \sum_{m=1}^{N} x_{k\ell m} (\alpha c_{k\ell} + \delta c_{\ell m})$$

$$\text{s.t.} \quad \sum_{k=1}^{N} r_{jk} = 1, \quad \forall j=1,\ldots,N \qquad (14)$$

$$\sum_{j=1}^{N} r_{jk} \le N y_k, \quad \forall k=1,\ldots,N \qquad (15)$$

$$r_{jj} = y_j, \quad \forall j=1,\ldots,N \qquad (16)$$

$$\sum_{\ell=1}^{N} x_{k\ell m} = \sum_{j=1}^{N} r_{jk} w_{jm}, \quad \forall k,m=1,\ldots,N \qquad (17)$$

$$\sum_{i=1}^{N} u_{ih} = \sum_{j=1}^{N} \sum_{\substack{k=1 \\ W_j c_{jk} \ge c_{(h)}}}^{N} r_{jk}, \quad \forall h=1,\ldots,G \qquad (18)$$

$$u_{ih} \ge u_{i-1h}, \quad \forall i=2,\ldots,N, \; h=1,\ldots,G \qquad (19)$$

$$\text{Constraints}: (8)-(11)$$
$$u_{ih}, r_{jk} \in \{0,1\}, \quad x_{k\ell m}, \; y_k \ge 0, \; \forall i,j,k,\ell,m=1,\ldots,N$$

Similarly to the 3-index formulation, in the above formulation, the family of constraints (8) and (16) are redundant whenever the cost structure satisfies the triangular inequality; however, they are useful in reducing solution times (see the Appendix for further details). Moreover, the family of constraints (15) can be presented in a disaggregated form, i.e.

$$r_{jk} \le y_k, \quad \forall j,k=1,\ldots,N$$

However, the computational experience provides worse computational running times with this alternative family of constraints. As in the previous section, variables y can be removed from this formulation using constraints (16), but we have kept them in order to clearly distinguish the location variables.

4. Comparing formulations

Before trying to improve the performance of the 3-index and covering 3-index formulations, we will compare them by means of a simple computational study. The formulations were implemented, as they have been presented in the previous sections, in the commercial solver Xpress IVE 1.19.01, running on a 2.40 GHz PC with 2.00 GB of RAM memory. The cut generation option of Xpress was disabled in order to compare the relative performance of the formulations cleanly.

We compare the performance of the two formulations presented in the previous section, namely 3-index and covering 3-index. In order to produce a set of test instances, we generated the data in two different ways. In our first set of instances we draw the unit costs and the origin–destination flows at random both in the interval [0,20], whereas the second set was also drawn at random the costs and the flows in [0,40]. This design tries to capture the difference between having more or less repeated data in the $(W_j c_{jk})$ matrix: the smaller the intervals the higher the number of repetitions. These two sets of data try to test whether the size of the parameter G, number of different cost values, is significant in order to solve the problem with the covering 3-index formulation. We tested the two formulations on a testbed of five instances for each combination of (i) costs matrices, (ii) N in {10,15,20}, (iii) different values of p depending on the case, and (iv) $\alpha=0.7$, $\delta=1.2\alpha$ and six different λ-vectors. Among them we consider p-center ($\lambda=(0,\dots,0,1)$, k-centrum ($k=\lceil 0.2N\rceil$, $\lambda=(0,\dots,0,1,\overset{k}{\dots},1)$), k_1+k_2-trimmed-mean ($k_1=k_2=\lceil 0.2N\rceil$, $\lambda=(0,\overset{k_1}{\dots},0,1,\dots,1,0,\overset{k_2}{\dots},0)$), anti-$k_1+k_2$-trimmed-mean ($k_1=k_2=\lceil 0.2N\rceil$, $\lambda=(1,\overset{k_1}{\dots},1,0,\dots,0,1,\overset{k_2}{\dots},1)$), p-median ($\lambda=(1,\dots,1)$) and non-increasing lambda weights ($\lambda=(\lambda_1,\dots,\lambda_N)$, with $\lambda_1\geq\dots\geq\lambda_N$).

Tables 2 and 3 report the results of the 3-index and covering 3-index formulations for these instances. The first column of these tables stands for the different types of problems in the study. Columns *Nodes*, *R-GAP*, *GAP* and *Time* stand for the averages of: number of nodes in the B&B tree, the gap in the root node, the final GAP (if any) after 1 h and the CPU time in seconds, respectively. (The time was limited to 1 h of CPU.) Values ">3600" of column *Time* means that Xpress requires more than 1 h of CPU time to solve each of the five instances for the corresponding combination of parameters whereas starred times (*) mean that at least one of the five instances was not solved to optimality within the time limit. In these cases, column GAP reports the gap at the stopping time.

In both tables we observe that formulation covering 3-index dominates in most cases formulation 3-index (with the exception of some cases with respect to the median problem). Moreover, the improvement of covering 3-index over 3-index increases as G decreases. This can be observed looking at the better behavior of

Table 2
First set of instances with costs and flows in $[0,20]\times[0,20]$.

N	***p***	**3-Index formulation**				**Covering 3-index formulation**			
		Nodes	**R-GAP**	**GAP**	**Time**	**Nodes**	**R-GAP**	**GAP**	**Time**
Center									
10	**3**	4710.20	28.98	0.00	11.14	1402.20	28.98	0.00	8.30
10	**5**	12394.60	22.81	0.00	26.84	812.60	22.81	0.00	4.81
15	**4**	326715.60	28.20	0.25	1723.59*	24945.20	28.20	0.00	282.52
15	**8**	391325.60	29.68	2.11	2191.61*	31162.00	29.37	0.00	245.48
20	**5**	197788.40	35.15	17.87	> 3600	158471.20	29.73	10.14	> 3600
20	**10**	159142.80	33.97	17.23	> 3600	212907.80	31.45	9.02	> 3600
k-Centrum									
10	**3**	6415.40	35.60	0.00	14.26	1261.40	35.60	0.00	6.73
10	**5**	2510.60	22.38	0.00	9.74	107.00	22.38	0.00	1.83
15	**4**	344490.20	36.74	0.38	1812.32	55281.00	36.74	0.00	450.75
15	**8**	69489.80	23.11	0.00	519.62	2761.00	27.73	0.00	44.35
20	**5**	183830.80	47.93	30.69	> 3600	176535.40	42.07	20.29	> 3600
20	**10**	128686.00	32.52	9.57	> 3600	90138.60	31.97	0.61	1791.55*
Median									
10	**3**	283.80	20.44	0.00	2.13	181.00	20.44	0.00	2.64
10	**5**	191.40	15.87	0.00	1.96	116.60	15.87	0.00	2.15
15	**4**	9384.80	28.24	0.00	91.02	22743.80	28.24	0.00	207.74
15	**8**	3741.80	23.11	0.00	49.65	2453.00	23.11	0.00	39.01
20	**5**	142778.80	40.95	15.03	> 3600	183695.80	34.40	9.17	> 3600
20	**10**	41495.40	27.01	0.00	1590.50	34151.80	27.01	0.00	675.22
Trimmed mean									
10	**3**	774.80	24.54	0.00	4.05	215.80	24.54	0.00	2.30
10	**5**	158.60	19.32	0.00	2.46	101.40	19.32	0.00	1.63
15	**4**	17999.40	29.70	0.00	117.30	4329.00	29.70	0.00	44.87
15	**8**	2238.60	25.42	0.00	39.01	1645.80	25.42	0.00	27.78
20	**5**	191894.00	35.85	4.89	> 3600	47821.40	33.83	0.00	785.20
20	**10**	45374.80	28.76	0.00	1634.72	13215.00	28.76	0.00	316.20
Anti-trimmed mean									
10	**3**	5555.80	32.21	0.00	11.45	1722.60	32.21	0.00	8.35
10	**5**	9261.40	29.01	0.00	19.83	562.20	29.01	0.00	3.41
15	**4**	552415.80	29.63	2.22	2746.05*	22274.00	29.02	0.00	231.25
15	**8**	401730.00	29.40	0.89	2375.21*	9063.00	29.39	0.00	93.37
20	**5**	184341.40	42.85	24.06	> 3600	128408.80	35.33	7.60	2901.55*
20	**10**	168067.20	34.98	16.75	> 3600	183725.60	34.31	6.38	> 3600
Non-increasing									
10	**3**	75.40	30.20	0.00	1.25	89.40	30.20	0.00	0.83
10	**5**	71.40	20.95	0.00	0.93	65.40	20.95	0.00	0.62
15	**4**	343.60	26.94	0.00	9.09	371.40	26.94	0.00	7.33
15	**8**	882.20	21.49	0.00	12.33	666.60	21.49	0.00	8.20
20	**5**	3225.00	32.59	0.00	137.71	3153.80	32.59	0.00	101.18
20	**10**	10139.80	27.36	0.00	274.99	8262.60	27.36	0.00	183.55

Table 3
Second set of instances with costs and flows in [0,40] × [0,40].

N	p	3-Index formulation				Covering 3-index formulation			
		Nodes	R-GAP	GAP	Time	Nodes	R-GAP	GAP	Time
Center									
10	**3**	5629.40	27.38	0.00	12.73	1106.40	27.35	0.00	7.19
10	**5**	9450.80	26.81	0.00	20.36	794.20	26.81	0.00	5.38
15	**4**	256355.20	34.45	0.48	1344.37	30681.60	34.45	0.00	339.48
15	**8**	459198.20	37.00	1.36	2688.13*	112346.40	36.98	0.00	743.97
20	**5**	185962.60	40.28	19.55	> 3600	121978.00	34.63	14.52	> 3600
20	**10**	165672.60	33.32	17.69	> 3600	178061.40	39.29	14.29	> 3600
k-Centrum									
10	**3**	4473.80	33.59	0.00	11.16	1021.40	33.59	0.00	5.65
10	**5**	1436.20	26.55	0.00	6.61	239.80	26.55	0.00	2.35
15	**4**	279503.00	41.55	0.00	1331.24	33547.20	41.55	0.00	319.93
15	**8**	68444.40	32.36	0.00	514.79	5528.60	32.36	0.00	80.50
20	**5**	191195.80	52.40	34.50	> 3600	153636.20	40.53	15.53	> 3600
20	**10**	123926.60	35.51	10.55	> 3600	86029.60	34.67	1.30	2136.19*
Median									
10	**3**	679.80	25.97	0.00	3.23	485.40	25.97	0.00	3.53
10	**5**	232.00	16.05	0.00	1.87	235.00	16.05	0.00	1.89
15	**4**	8573.80	31.23	0.00	74.27	27922.00	31.23	0.00	258.37
15	**8**	2955.40	26.64	0.00	37.50	2450.60	26.64	0.00	40.45
20	**5**	127737.20	36.26	9.24	3089.65*	126135.60	35.14	7.37	3006.37*
20	**10**	23267.20	28.49	0.00	960.57	52692.00	28.49	0.00	880.28
Trimmed mean									
10	**3**	727.40	29.38	0.01	3.94	252.40	29.38	0.00	2.77
10	**5**	139.00	18.88	0.01	2.57	98.20	18.88	0.00	1.93
15	**4**	17202.00	32.50	0.00	110.28	10487.20	32.50	0.00	100.20
15	**8**	1713.80	26.55	0.00	33.49	1085.80	26.55	0.00	22.87
20	**5**	229109.40	36.08	1.33	> 3600	121019.20	35.82	2.11	2301.74
20	**10**	30722.00	29.68	0.00	1109.77	16320.00	29.68	0.00	455.58
Anti-trimmed mean									
10	**3**	5445.20	36.13	0.00	11.78	1220.60	36.13	0.00	7.24
10	**5**	8580.20	34.42	0.00	18.28	763.80	34.42	0.00	4.56
15	**4**	328888.74	42.50	1.55	2069.58*	39290.40	42.25	0.00	380.00
15	**8**	447332.60	37.83	1.57	2697.39*	15652.40	37.83	0.00	139.11
20	**5**	199909.20	45.19	25.36	> 3600	129636.40	38.04	12.30	> 3600
20	**10**	165004.80	38.72	19.91	> 3600	179333.60	37.23	11.09	> 3600
Non-increasing									
10	**3**	48.60	26.16	0.00	1.01	46.60	26.16	0.00	0.67
10	**5**	41.80	19.82	0.00	1.28	41.40	19.82	0.00	0.61
15	**4**	668.60	37.08	0.00	13.69	637.80	37.08	0.00	10.51
15	**8**	881.40	30.05	0.00	14.27	795.00	30.05	0.00	10.30
20	**5**	4867.00	35.58	0.00	202.48	5525.40	35.58	0.00	171.03
20	**10**	13831.80	31.21	0.00	366.12	11827.00	31.21	0.00	302.44

covering 3-index on the instances of Table 2 where the number of repetitions is greater than in the instances of Table 3 and consequently, the value of G is lower. This fact reinforces our claim that covering 3-index is a better formulation as G decreases; recall that the number of variables and constraints of this formulation depends on G. (The reader can check that the number of integer variables and constraints of formulation 3-index is N^3 and N^3+N^2+7N+1, respectively, whereas the same figures for the covering 3-index formulation are N^2+NG and $N^3+N^2+(N+1)G+5N+1$, see Table 6.) It is also of interest to look at the gap at the root node. Both formulations provide similar gaps at the root node (therefore their linear relaxation is almost identical).

Moreover, Tables 2 and 3 show, as a general trend, that the CPU time increases similarly, for all choices of the λ-vector, with the size of the instances. Formulation 3-index proves to be unable to reach optimality for instances of size $N=20$, except for the cases of the median problem ($p=10$) and λ-weights given in non-increasing order. This latter choice of lambdas resulted in a specially easy problem. This effect may be due to the fact that the components of the λ-weights are different, thus avoiding degeneracy. A similar behavior is observed with regard to formulation covering 3-index, although in this case the resulting gaps are significantly smaller.

We have also done the same analysis for $\alpha=0.2$ and $\delta=0.84$, obtaining similar results, although, in general, this model needed more time to be solved. Moreover, since in this case the inter-hub transportation costs have a larger discount, the number of origin–destination paths crossing two different hubs has increased around 8%.

5. Improvements

This section is devoted to present some improvements and strengthenings of the previous formulations as well as an alternative formulation, only valid for non-decreasing lambdas, that in some cases give good results in order to reduce the CPU time. The goal pursued with these new approaches is to reduce the CPU times needed to solve the problems with our earlier formulations.

5.1. An alternative formulation: OT-3-index

This section describes an alternative formulation, that is only valid under the hypothesis that λ-weights are in non-decreasing order, that is, $0 := \lambda_0 \le \lambda_1 \le \cdots \le \lambda_N$. Although, it is more restrictive than the general formulations given above it covers most of the classical problems in the literature of hub location and moreover it is rather efficient in order to speed up the computing times. The rationale of this approach is based on the work by Ogryczak and Tamir [33]. We have adapted their formulation for the ordered median problem to our case as follows.

Taking into account the variables r_{jk} defined in the covering 3-index formulation, we define the variables $v_j = W_j \sum_{k=1}^{N} c_{jk} r_{jk}$. Following the idea of Ogryczak and Tamir, we look for the sum of the q largest v-values. $v_{(i)}$ represents the values v_i sorted in non-increasing order.

For any integer q, $1 \le q \le N$, consider the following function defined in $[0,+\infty)$:

$$f_q(z) := qz + \sum_{k=1}^{N}\sum_{i=1}^{N} \max\{0, v_j - z\}$$

This is a convex piecewise linear function with slopes moving from $q-N$ to q in integer steps, whose minimum is reached either when the slope is 0 or, if this is not the case, when the slope changes from negative to positive, i.e., when z equals the qth maximum value of the vector v, namely $v_{(q)}$. Thus, the minimum value of f is

$$f_q(v_{(q)}) = qv_{(q)} + \sum_{j=1}^{N} \max\{0, v_j - v_{(q)}\} = qv_{(q)} + \sum_{j=1}^{q}(v_{(j)} - v_{(q)}) = \sum_{j=1}^{q} v_{(j)}$$

i.e., the sum of the q maximum values. Therefore, minimizing the sum of the q-largest v-values can be linearized as

$$\min\ qz_q + \sum_{i=1}^{N} D_{iq} \quad \text{s.t. } D_{iq} \ge 0\ \ \forall i, \quad D_{iq} \ge v_j - z_q\ \ \forall j = 1,\ldots,N$$

In the following we add the index q in variables D and z because we are interested in solving this problem for each value q from 1 to $N-1$ and combine the latter linearization with the constraints used in the former formulations to get the values of the v_j variables in our formulations.

Therefore, taking into account the above discussion, the OT-3-index formulation is

$$\begin{aligned}
\min\quad & \sum_{i=1}^{N}(\lambda_{N-i+1} - \lambda_{N-i})\left(iz_i + \sum_{j=1}^{N}\sum_{j=1}^{N} D_{ij}\right) \\
& + \sum_{k=1}^{N}\sum_{\ell=1}^{N}\sum_{m=1}^{N}(\alpha c_{k\ell} + \delta c_{\ell m}) x_{k\ell m} \\
\text{s.t.}\quad & D_{ij} \ge \sum_{k=1}^{N}\sum_{m=1}^{N} W_j c_{jk} r_{jk} - z_i, \quad \forall i,j = 1,\ldots,N \\
& D_{ij} \ge 0, \quad \forall i,j = 1,\ldots,N \\
& z_i, \text{ unrestricted}, \quad \forall i = 1,\ldots,N \\
& \text{Constraints}: (8)-(11),\ (14)-(17) \\
& r_{jk}, y_k \in \{0,1\}, \quad x_{k\ell m} \ge 0,\ \forall i,j,k,\ell,m = 1,\ldots,N
\end{aligned}$$

5.2. Variable fixing

This section addresses the description of some preprocessing steps that we propose to reduce the size of our covering 3-index formulation. Due to the definition of the variables in covering 3-index formulation, one can expect that many u-variables in the right-hand part of the matrix of u-variables will take value 0 in the optimal solution. Indeed, $u_{ih}=0$ means that the ith sorted allocation cost is less than or equal to $c_{(h)}$ which is very likely to be true if h is sufficiently large and i is perhaps not that large. The same type of arguments also suggest that one may expect that $u_{ih}=1$ whenever i is large and h is small to medium size because this would mean that the ith sorted allocation cost would not have been done at cost less than $c_{(h)}$. With these strategies, the size of the formulation could be reduced if some (hopefully many) of these variables were fixed beforehand. In this subsection we describe a number of variable fixing possibilities for the set of u-variables which are useful in the overall solution process.

First of all, it is clear that since $c_{jj}=0, \forall j = 1,\ldots,N$ we have that

$$u_{i1} = 1, \quad \forall i = 1,\ldots,N$$

$$u_{i2} = 0, \quad \forall i = 1,\ldots,p \tag{20}$$

Moreover, whenever $W_j c_{jk} \ne 0$ if and only if $j \ne k$ then we can also fix

$$u_{i2} = 1, \quad \forall i = p+1,\ldots,N$$

5.2.1. Fixing u-variables to 1

In order to fix u_{ih}-variables to 1 for a given $h \in \{1,\ldots,G\}$, we will deal with an auxiliary problem that maximizes the number of origin-first hub allocations satisfying $W_j c_{jk} \le c_{(h-1)}$ which is equivalent to the maximum number of variables u_{ih} that can assume a zero value. Let

$$z_{jk} = \begin{cases} 1 & \text{if origin site } j \text{ is assigned to hub } k \\ 0 & \text{otherwise} \end{cases}$$

Using these variables, the formulation of this problem is

$$\begin{aligned}
\max\quad & H1_h := \sum_{j=1}^{N}\sum_{k=1}^{N} z_{jk} \\
\text{s.t.}\quad & z_{jk} c_{jk} \sum_{m=1}^{N} w_{jm} \le c_{(h-1)}, \quad \forall j,k = 1,\ldots,N \\
& \sum_{k=1}^{N} z_{jk} \le 1, \quad \forall j = 1,\ldots,N \\
& z_{jk} \le y_k, \quad \forall j,k = 1,\ldots,N \\
& \sum_{k=1}^{N} y_k \le p \\
& z_{jk}, y_k \in \{0,1\}, \quad j,k,m = 1,\ldots,N
\end{aligned}$$

If $H\ 1_h$ is the optimal value of problem above, since there are N origin-first hub allocations, the number of allocations satisfying $W_j c_{jk} > c_{(h)}$ must be necessarily greater than or equal to $N - H1_h + 1$, or equivalently, in any feasible solution of covering 3-index formulations

$$u_{ih} = 1, \quad \forall i = H1_h,\ldots,N$$

5.2.2. Fixing u-variables to 0

Under a similar rationale to the one used in the previous section, we try to fix as many u_{ih}-variables to 0 as possible, for a given $h \in \{1,\ldots,G\}$. In this case, we deal with an auxiliary problem that maximizes the number of origin-first hub allocations satisfying $W_j c_{jk} \ge c_{(h)}$. In conclusion, this auxiliary problem provides the minimum number of zeros that the hth column of the u-matrix must have

$$\max\quad H2_h := \sum_{j=1}^{N}\sum_{k=1}^{N} z_{jk}$$

$$\text{s.t.}\quad c_{jk} \sum_{m=1}^{N} w_{jm} \ge z_{jk} c_{(h)}, \quad \forall j,k = 1,\ldots,N$$

$$\sum_{k=1}^{N} z_{jk} \le 1, \quad \forall j = 1, \ldots, N$$
$$z_{jk} \le y_k, \quad \forall j,k = 1, \ldots, N$$
$$\sum_{k=1}^{N} y_k \le p$$
$$y_k, z_{jk} \in \{0,1\}, \quad \forall j,k = 1, \ldots, N$$

Therefore, if $H2_h$ is the optimal value of problem above, the hth column of the u-matrix must have at least $N-H2_h$ zeros, i.e., in any feasible solution of the covering 3-index formulations:

$$u_{ih} = 0, \quad \forall i = 1, \ldots, N - H2_h$$

5.3. Valid inequalities

Next, we discuss some valid inequalities that strengthen the covering 3-index formulation. First, we present a family of valid inequalities that are a straightforward consequence of the definition of the u-variables (see (13)), but that help a lot in solving the problem (see Tables 4 and 5):

$$u_{ih} \ge u_{ih+1}, \quad h = 1, \ldots, G-1 \tag{21}$$

Note that (21) states that if an allocation is done at cost $c_{(h)}$, then it will not be done at larger cost. Based on this set of inequalities and using (20), we can also fix the following variables to 0:

$$u_{ih} = 0, \quad \forall i = 1, \ldots, p, \; h = 2, \ldots, G \tag{22}$$

The last family of valid inequalities state disjunctive implications on the origin-first hub allocation costs. The first one ensures the following implications: (1) if there are no hubs open to allocate origin site j at a cost less than $c_{(h)}$ then origin site j must be allocated to a first hub at a cost at least $c_{(h)}$, and (2) if origin site j is allocated at a cost less than $c_{(h)}$ then at least one open hub must satisfy $W_j c_{jk} < c_{(h)}$:

$$\sum_{k=1: W_j c_{jk} \ge c_{(h)}}^{N} r_{jk} + \sum_{k=1: W_j c_{jk} < c_{(h)}}^{N} y_k \ge 1, \quad \forall j = 1, \ldots, N, \; h = 1, \ldots, G \tag{23}$$

Swapping the roles of the r- and y-variables in the above inequality, we obtain the following complementary

Table 4
First set of instances with costs and flows in $[0,20] \times [0,20]$.

N	p	OT-3-index				Pre-covering 3-index				
		Nodes	R-GAP	GAP	Time	Nodes	R-GAP	GAP	Time	Fixed var.
Center										
10	**3**	73.00	22.29	0.00	0.41	220.20	21.97	0.00	3.20	39.94
10	**5**	102.60	17.02	0.00	0.43	176.60	15.00	0.00	2.22	48.59
15	**4**	771.40	24.84	0.00	6.75	2324.80	25.27	0.00	67.01	33.60
15	**8**	1291.40	25.49	0.00	8.11	2797.80	25.19	0.00	41.68	51.45
20	**5**	4162.20	26.32	0.00	69.52	12074.00	27.26	0.00	770.49	31.51
20	**10**	11323.60	28.41	0.00	139.44	26828.80	28.58	0.00	858.51	49.01
k-Centrum										
10	**3**	164.20	31.40	0.00	0.39	196.60	25.46	0.00	2.16	38.65
10	**5**	61.00	19.66	0.00	0.32	63.00	15.08	0.00	0.91	48.76
15	**4**	651.40	33.44	0.00	5.40	1966.60	28.89	0.00	47.84	34.42
15	**8**	692.20	25.27	0.00	4.82	768.20	21.88	0.00	13.19	51.07
20	**5**	7362.20	39.86	0.00	108.15	30714.00	36.09	0.00	1501.95	31.26
20	**10**	8593.80	29.55	0.00	110.90	7767.20	27.53	0.00	237.74	48.58
Median										
10	**3**	32.60	20.44	0.00	0.25	59.00	18.91	0.00	1.16	37.11
10	**5**	43.00	15.87	0.00	0.26	55.80	15.66	0.00	1.05	48.44
15	**4**	753.40	28.24	0.00	5.62	1624.40	27.05	0.00	33.15	33.68
15	**8**	775.40	23.11	0.00	5.27	1080.20	22.93	0.00	18.07	51.42
20	**5**	6301.40	33.60	0.00	84.94	12204.40	31.81	0.00	581.18	25.83
20	**10**	11705.80	27.01	0.00	142.53	13672.20	26.95	0.00	378.88	48.81
Trimmed mean										
10	**3**					85.80	20.24	0.00	1.18	37.65
10	**5**					58.20	16.20	0.00	0.95	48.60
15	**4**					1024.20	25.87	0.00	22.76	33.48
15	**8**					772.60	23.47	0.00	13.85	51.55
20	**5**					5444.00	30.33	0.00	236.80	30.95
20	**10**					7022.60	27.08	0.00	207.12	48.68
Anti-trimmed mean										
10	**3**					328.60	27.05	0.00	3.53	37.35
10	**5**					207.40	22.96	0.00	1.98	48.55
15	**4**					2816.20	24.96	0.00	67.72	33.75
15	**8**					2733.40	24.56	0.00	36.49	51.03
20	**5**					31927.60	31.13	0.00	1546.92	31.52
20	**10**					46558.60	30.56	0.00	1376.21	48.87
Non-increasing										
10	**3**					86.20	27.98	0.00	0.77	37.27
10	**5**					58.60	20.65	0.00	0.60	47.73
15	**4**					335.80	25.92	0.00	7.24	33.01
15	**8**					676.60	21.40	0.00	8.67	51.54
20	**5**					3437.40	31.44	0.00	110.97	31.22
20	**10**					8733.40	27.36	0.00	196.60	48.63

Table 5
Second set of instances with costs and flows in [0,40] × [0,40].

N	p	OT-3-index				Pre-covering 3-index				
		Nodes	R-GAP	GAP	Time	Nodes	R-GAP	GAP	Time	Fixed var.
Center										
10	3	57.00	22.25	0.00	0.39	218.60	20.85	0.00	3.46	38.99
10	5	81.40	20.94	0.00	0.41	147.00	20.24	0.00	2.31	49.15
15	4	444.20	29.96	0.00	4.86	1362.20	31.08	0.00	59.83	33.82
15	8	2165.00	33.42	0.00	12.67	4058.80	33.18	0.00	66.98	51.53
20	5	2329.80	29.73	0.00	45.05	10140.40	30.96	0.00	845.07	31.17
20	10	119922.00	35.29	5.78	2276.80	30918.20	35.56	0.00	1164.23	48.97
k-Centrum										
10	3	55.00	30.30	0.00	0.40	82.60	25.71	0.00	1.72	37.80
10	5	66.60	22.42	0.00	0.37	82.60	19.33	0.00	1.01	49.51
15	4	700.20	38.23	0.00	5.70	2182.40	35.27	0.00	73.92	33.44
15	8	742.20	29.44	0.00	5.49	820.20	26.87	0.00	16.72	51.69
20	5	3939.80	36.97	0.00	67.88	18786.40	33.77	0.00	1126.53	31.33
20	10	43973.40	32.71	1.01	823.28	10091.20	30.52	0.00	317.69	49.15
Median										
10	3	65.00	25.97	0.00	0.31	99.00	23.36	0.00	1.63	37.93
10	5	55.80	16.05	0.00	0.29	69.80	15.98	0.00	0.99	49.17
15	4	449.00	31.23	0.00	3.88	1439.00	29.47	0.00	36.69	34.48
15	8	849.40	26.64	0.00	5.58	936.60	26.63	0.00	18.76	52.35
20	5	8691.00	33.32	0.00	142.01	13151.60	32.62	0.00	729.93	31.35
20	10	6856.60	28.49	0.00	84.95	9276.00	28.48	0.00	282.83	49.20
Trimmed mean										
10	3					80.20	25.51	0.00	1.48	37.34
10	5					64.20	16.86	0.00	1.02	48.88
15	4					719.40	29.09	0.00	20.45	33.82
15	8					581.00	24.83	0.00	12.82	51.95
20	5					5533.80	32.80	0.00	283.45	31.16
20	10					7870.60	28.44	0.00	272.70	49.24
Anti-trimmed mean										
10	3					545.40	29.67	0.00	4.70	38.44
10	5					265.00	28.49	0.00	2.32	48.91
15	4					2856.20	38.01	0.00	79.01	33.46
15	8					3555.00	33.43	0.00	50.04	51.98
20	5					22297.80	34.10	0.00	1255.70	31.42
20	10					56598.20	33.28	0.00	1551.59	49.45
Non-increasing										
10	3					69.00	24.09	0.00	0.81	38.51
10	5					52.20	19.57	0.00	0.62	49.25
15	4					621.40	35.29	0.00	11.05	34.29
15	8					943.80	30.05	0.00	11.69	52.18
20	5					4789.80	33.68	0.00	162.28	31.08
20	10					13228.40	31.21	0.00	307.14	49.33

inequalities:

$$\sum_{k=1:W_jc_{jk}\geq c_{(h)}}^{N} y_k + \sum_{k=1:W_jc_{jk}<c_{(h)}}^{N} r_{jk} \geq 1, \quad \forall j=1,\ldots,N,\ h=1,\ldots,G \tag{24}$$

The families of valid inequalities (21)–(24) have been included, together with the preprocessing for variable fixing presented in Section 5.2, under the name of pre-covering 3-index in order to be compared with the previous formulations. The results are reported in Section 6.

6. Comparing improved formulations

This section reports on the computational comparison between pre-covering 3-index, the strengthening of formulation covering 3-index given by the preprocessing phase described in Section 5.2 and the valid inequalities of Section 5.3, and OT-3-index formulations. For this analysis we follow the same pattern as in Section 4. We consider the same two sets of instances with costs and flows randomly generated in [0,20] and [0,40] and the same six families of problem types: center, k-centrum, median, k_1+k_2-trimmed-mean, anti-k_1+k_2-trimmed-mean and non-increasing λ-weights. As for the presentation of results we also follow the same structure as in Section 4. Tables 4 and 5 report the results for first and second sets of instances, respectively. Each table has three blocks of columns. The first block includes the name of the problem and the sizes of the instances (N and p). Then, the next two blocks show the results obtained by the OT-3-index and pre-covering 3-index formulations. In both blocks, we report the same information as in Tables 2 and 3, i.e. the averages of: number of nodes of the B&B tree (*Nodes*), gaps a the root node (*R-GAP*), final gaps (*GAP*) and CPU times in seconds (*Time*). In addition, in the third block we report the percentage of integer variables that are fixed by our preprocessing on the pre-covering 3-index formulation (*Fixed var.*). Note that the empty blocks in Tables 4 and 5 appear because OT-3-index is not applicable to the corresponding problems (the components of the λ-vector are not given in non-decreasing order).

We observe that both formulations are more efficient than the most efficient one considered previously, namely covering

Table 6
Comparison of different formulations.

	Constraints	Continuous variables	Integer variables
3-Index formulation	N^3+N^2+7N+1	$2N^3$	N^3
COV-3-index formulation	$N^3+N^2+(N+1)G+5N+1$	N^3+N^2+NG	N^2+NG
OT-3-index formulation	N^3+2N^2+5N+1	N^3+N^2+N	N^2+N

3-index, in almost all cases. The only exceptions appear in the problems with non-increasing lambdas, although in these cases the CPU are almost similar. This can be seen by comparing the CPU times of Tables 2 and 4 for the first set of instances and Tables 3 and 5 for the second set of instances. We also remark the efficiency of our preprocessing, which is illustrated by the number of fixed integer variables.

Moreover, the OT-3-index formulation (recall that this model is only applicable for non-decreasing λ-weights) gets extremely good results for the first set of instances (Table 4), in fact, the maximum average time needed to solve a battery of five instances of any type of analyzed problem is 142.53 s (median problem with $N=20$, $p=10$). The formulation pre-covering 3-index is not as good, although it can be applied to any configuration of λ-weights. Nevertheless, the comparison is not as neat on the second set of instances (see Table 5). In this case, OT-3-index solves easily all instances up to size $N=15$ but it fails to solve several instances for sizes $N=20$, $p=10$ (3 out of 5 of the center and 1 out of 5 of the k-centrum) whereas pre-covering 3-index solves all instances. Concerning the gap at the root node there are no significative difference between both formulations in any of the two set of instances. We have also tried larger instances. In these cases, already for $N=23$, the number of instances that runs out of memory increases rapidly. It is remarkable to report that for $N=26$ all tested instances run out of memory in both formulations. In any case, both formulations improve the behavior of covering 3-index.

Finally, in Table 6, we provide a summary of the number of constraints, number of continuous and integer variables of the different formulations that we have proposed in this paper. 3-Index formulation has the fewest number of constraints but it is the formulation with the largest number of continuous and integer variables. On the opposite side, OT-3-index formulation has the largest number of constraints but it is the formulation with the fewest continuous and integer variables.

7. Conclusions

This paper can be considered as an attempt to deal with new flexible formulations for hub location problems. Although, we focus mainly on modeling issues, we also report computational tests comparing the performance of the different formulations. These results aim to establish the limits, both on CPU-time and size, of the exact resolution of the different models using standard MIP-solvers. Our preliminary analysis shows that "ad hoc" methods are required to solve even small size instances of these problems. Thus, this paper provides a starting point for the development of exact and heuristic solution methods for all the models that have been introduced.

Appendix A

In this section, we present a structural result about the linear programming representation of the feasible set of our problem. It states that some constraints, that in the general case are required to get a valid formulation, are redundant whenever the cost structure satisfies the triangular inequality.

Proposition A.1. *If the cost structure satisfies the triangular inequality, $\alpha \le \lambda_i$ for any $i=1,\ldots,N$ and $\alpha \le \delta$ then*

(i) *Inequality* (5) *is redundant.*
(ii) *Inequality* (8) *is redundant.*

Proof. We distinguish two cases. If $y_j=0$ the inequality (5) follows from (4). In the case where $y_j=1$, we argue by contradiction to prove that $\sum_{i=1}^N r^i_{jj}=1$. If $r^i_{jj}=0$ for all $i=1,\ldots,N$ then by (2) there is an index $k^*(\neq j)$, such that, $r^i_{jk^*}=1$ for some $i=1,\ldots,N$. For any destination site m, the flow (j,m) should have one of the following patterns: $(j,k^*,\ell^*(m),m)$ or (j,k^*,k^*,m), for some $k^*,\ell^*(m)$ sites in $\{1,\ldots,N\}$ that are open hubs (observe that for the case $y_m=1$, we have that $\ell^*(m)=m$). The first one stands for a path going through two different hubs $(k^*,\ell^*(m))$, whereas the second one means going via a single hub k^*. Therefore, this solution should be cheaper than the one with path $(j,j,\ell(m),m)$ and (j,j,j,m) for any open hub $\ell(m)$. Therefore, the following inequality would be satisfied:

$$\min\left\{\lambda_i c_{jk^*}W_j+\alpha\sum_{m=1}^N c_{k^*,\ell^*(m)}w_{jm} +\delta\sum_{m=1}^N c_{\ell^*(m),m}w_{jm},\lambda_i c_{jk^*}W_j+\delta\sum_{m=1}^N c_{k^*m}w_{jm}\right\} < \min\left\{\min_{\ell(m)}\left\{0+\alpha\sum_{m=1}^N c_{j,\ell(m)}w_{jm}. +\delta\sum_{m=1}^N c_{\ell(m),m}w_{jm}\right\},0+0+\delta\sum_{m=1}^N c_{jm}w_{jm}\right\}$$

However, since the triangular inequalities hold and $\alpha\le\lambda_i$ for any $i=1,\ldots,N$, we obtain that

$$\min_{\ell(m)}\left\{\alpha\sum_{m=1}^N c_{j,\ell(m)}w_{jm}+\delta\sum_{m=1}^N c_{\ell(m),m}w_{jm}\right\} \le \alpha\sum_{m=1}^N c_{j,\ell^*(m)}w_{jm}+\delta\sum_{m=1}^N c_{\ell^*(m),m}w_{jm} \le \lambda_i c_{jk^*}W_j+\alpha\sum_{m=1}^N c_{k^*,\ell^*(m)}w_{jm}+\delta\sum_{m=1}^N c_{\ell^*(m),m}w_{jm}$$

and

$$\min_{\ell(m)}\left\{\alpha\sum_{m=1}^N c_{j,\ell(m)}w_{jm}+\delta\sum_{m=1}^N c_{\ell(m),m}w_{jm}\right\} \le \alpha c_{jk^*}W_j+\delta\sum_{m=1}^N c_{k^*m}w_{jm}\le\lambda_i c_{jk^*}W_j+\delta\sum_{m=1}^N c_{k^*m}w_{jm}$$

The proof of the second assertion is similar to the first one and therefore it is omitted. □

Table 7
Applying Proposition A.1.

N	p	Covering 3 index				Covering 3-index Proposition A.1			
		Nodes	R-GAP	GAP	Time	Nodes	R-GAP	GAP	Time
10	**3**	350.00	15.71	0.00	2.28	718.20	33.95	0.00	2.98
10	**5**	146.40	10.32	0.00	1.19	666.60	22.47	0.00	2.28
15	**4**	5357.80	14.79	0.00	56.73	109670.20	31.96	1.92	1005.76*
15	**8**	147.00	6.59	0.00	6.09	12921.80	13.61	0.00	76.63
20	**5**	46315.20	15.16	1.12	1643.18*	195142.80	31.90	15.56	> 3600
20	**10**	2810.00	5.89	0.00	112.17	282035.40	15.89	6.79	> 3600

Similarly to the 3-index formulation, Proposition A.1 is also valid for the covering 3-index formulation applied to constraints (8) and (16). In the case that we are under the conditions of Proposition A.1, despite that constraints (5) and (8), as well as, (8) and (16) are redundant for the 3-index and covering 3-index formulations, respectively, our computational experience shows that it is much better to add them to the model, as valid inequalities, to reduce CPU time in solving the problem. Indeed, Table 7 reports on the average of five instances of different combinations of N in {10,15,20}, p depending on the value N and cost structure satisfying the triangular inequality (costs proportional to ℓ_1 distances between randomly generated integer coordinate points in $[0,50]\times[0,50]$), for a set of lambdas generated randomly in $[\delta+1,\delta+2]$, $\delta=1.2\alpha$ and $\alpha=0.7$. Looking at Table 7, we first observe that the R-GAP column, the one that gives the value of the linear relaxation of the problem, is much better adding inequalities (8) and (16), and reducing the gap at the root node almost by one-half. Moreover, the average number of nodes in the B&B tree is considerably lower for the formulation with constraints (8) and (16). Thus, one can conclude that in all cases adding these constraints helps in solving the problems. Therefore, although both set of constraints are redundant, provided that the triangular inequality holds, it is useful to consider them as valid inequalities in any resolution scheme. It is worth mentioning that the same behavior, concerning the inclusion of the two groups of constraints (5) and (8), is also observed under the 3-index formulation although we do not include the computational results for the sake of brevity.

References

[1] Alumur S, Kara BY. Network hub location problems: the state of the art. European Journal of Operational Research 2008;190(1):1–21.
[2] Berman O, Kalcsics J, Krass D, Nickel S. The ordered gradual covering location problem on a network. Discrete Applied Mathematics 2009;157(18–28): 3689–707.
[3] Boland N, Domínguez-Marín P, Nickel S, Puerto J. Exact procedures for solving the discrete ordered median problem. Computers and Operations Research 2006;33:3270–300.
[4] Boland N, Krishnamoorthy M, Ernst AT, Ebery J. Preprocessing and cutting for multiple allocation hub location problems. European Journal of Operational Research 2004;155(3):638–53.
[5] Bollapragada R, Li Y, Rao US. Budget-constrained, capacitated hub location to maximize expected demand coverage in fixed-wireless telecommunication networks. INFORMS Journal on Computing 2006;18(4):422–32.
[6] Cánovas L, García S, Labbé M, Marín A. A strengthened formulation for the simple plant location problem with order. Operations Research Letters 2007;35(2):141–50.
[7] Cánovas L, García S, Marín A. Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. European Journal of Operational Research 2007;179(3):990–1007.
[8] Campbell JF. Hub location and the p-hub median problem. Operations Research 1996;44(6):923–35.
[9] Campbell JF, Ernst A, Krishnamoorthy M. Hub location problems. In: Facility location. Berlin: Springer; 2002. p. 373–407.
[10] Campbell AM, Lowe TJ, Zhang L. The p-hub center allocation problem. European Journal of Operational Research 2007;176(2):819–35.
[11] Contreras I, Díaz JA, Fernández E. Lagrangian relaxation for the capacitated hub location problem with single assignment. OR Spectrum 2009;31(3): 483–505.
[12] Elloumi S, Labbé M, Pochet Y. A new formulation and resolution method for the p-center problem. INFORMS Journal on Computing 2004;16(1):84–94.
[13] Ernst AT, Krishnamoorthy M. Efficient algorithms for the uncapacitated single allocation p-hub median problem. Location Science 1996;4(3):139–54.
[14] Ernst AT, Krishnamoorthy M. Solution algorithms for the capacitated single allocation hub location problem. Annals of Operations Research 1999;86: 141–59.
[15] Espejo I, Marín A, Puerto J, Rodríguez-Chía AM. A comparison of formulations and solution methods for the minimum-envy location problem. Computers and Operations Research 2009;36:1966–81.
[16] Fonseca MC, García-Sánchez A, Ortega-Mier M, Saldanha-da-Gama F. A stochastic bi-objective location model for strategic reverse logistics. TOP 2010;18(1):158–84.
[17] Hamacher H, Labbé M, Nickel S, Sonneborn T. Adapting polyhedral properties from facility to hub location problems. Discrete Applied Mathematics 2004;145(1):104–16.
[18] Kalcsics J, Nickel S, Puerto J, Rodríguez-Chía AM. Distribution systems design with role dependent objectives. European Journal of Operational Research 2010;202:491–501.
[19] Kalcsics J, Nickel S, Puerto J, Rodríguez-Chía AM. The ordered capacitated facility location problem. TOP 2009;18(1):203–22.
[20] Kara BY, Tansel BC. On the single-assignment p-hub center problem. European Journal of Operational Research 2000;125(3):648–55.
[21] Kara BY, Tansel BC. The single-assignment hub covering problem: models and linearizations. Journal of the Operational Research Society 2003;54(1):59–64.
[22] Kolen A. Solving covering problems and the uncapacitated plant location problem on trees. European Journal of Operational Research 1983;12(3): 266–78.
[23] Kratica J, Stanimirovic Z. Solving the uncapacitated multiple allocation p-hub center problem by genetic algorithm. Asia-Pacific Journal of Operational Research 2006;23(4):425–37.
[24] Labbé M, Yaman H. Projecting the flow variables for hub location problems. Networks 2004;44(2):84–93.
[25] Labbé M, Yaman H. Solving the hub location problem in a star-star network. Networks 2008;51(1):19–33.
[26] Labbé M, Yaman H, Gourdin E. A branch and cut algorithm for hub location problems with single assignment. Mathematical Programming 2005;102(2): 371–405.
[27] Marín A. Formulating and solving splittable capacitated multiple allocation hub location problems. Computers and Operations Research 2005;32(12): 3093–109.
[28] Marín A. Uncapacitated Euclidean hub location: strengthened formulation, new facets and a relax-and-cut algorithm. Journal of Global Optimization 2005;33(3):393–422.
[29] Marín A, Cánovas L, Landete M. New formulations for the uncapacitated multiple allocation hub location problem. European Journal of Operational Research 2006;172(1):274–92.
[30] Marín A, Nickel S, Puerto J, Velten S. A flexible model and efficient solution strategies for discrete location problems. Discrete Applied Mathematics 2009;157(5):1128–45.
[31] Meyer T, Ernst AT, Krishnamoorthy M. A 2-phase algorithm for solving the single allocation p-hub center problem. Computers and Operations Research 2009;36(12):3143–51.
[32] Nickel S, Puerto J. Location theory—a unified approach. Springer; 2005.
[33] Ogryczak W, Tamir A. Minimizing the sum of the k largest functions in linear time. Information Processing Letters 2003;85:117–22.
[34] O'Kelly ME. A quadratic integer problem for the location of interacting hub facilities. European Journal of Operational Research 1987;32:393–404.
[35] Puerto J, Fernández FR. Geometrical properties of the symmetrical single facility location problem. Journal of Nonlinear and Convex Analysis 2000;1(3):321–42.

[36] Rodríguez-Chía AM, Nickel S, Puerto J, Fernández FR. A flexible approach to location problems. Mathematical Methods of Operations Research 2000;51: 69–89.
[37] Rodríguez-Martín I, Salazar-González JJ. Solving a capacitated hub location problem. European Journal of Operational Research 2008;184(2): 468–79.
[38] Tan PZ, Kara BY. A hub covering model for cargo delivery systems. Networks 2007;49(1):28–39.
[39] Wagner B. Model formulations for hub covering problems. Journal of the Operational Research Society 2008;59(7):932–8.
[40] Yaman H. Polyhedral analysis for the uncapacitated hub location problem with modular arc capacities. SIAM Journal on Discrete Mathematics 2005; 19(2):501–22.
[41] Zhou G, Min H, Gen M. The balanced allocation of customers to multiple distribution centers in a supply chain network: a genetic algorithm approach. Computers and Industrial Engineering 2002;43:251–61.