

UPCommons

Portal del coneixement obert de la UPC

<http://upcommons.upc.edu/e-prints>

Aquesta és una còpia de la versió *author's final draft* d'un article publicat a la revista *Computer and Operations Research*.

URL d'aquest document a UPCommons E-prints:

<http://hdl.handle.net/2117/134363>

Article publicat / Published paper:

Albareda-Sambola, M.; Martínez-Merino, L.; Rodríguez-Chía, A.M. The stratified p-center problem. *Computers & operations research*, 1 Agost 2019, vol. 108, p. 213-225. DOI: <[10.1016/j.cor.2019.04.013](https://doi.org/10.1016/j.cor.2019.04.013)>.

© <2019>. Aquesta versió està disponible sota la llicència CC-BY- NC-ND 4.0 <http://creativecommons.org/licenses/by-nc-nd/4.0/>

The stratified p -center problem

Maria Albareda-Sambola¹, Luisa I. Martínez-Merino², Antonio M. Rodríguez-Chía²

¹ Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya-BarcelonaTech, Terrassa, Spain

²Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain

Abstract

This work presents an extension of the discrete p -center problem. In this new model, called Stratified p -Center Problem ($SpCP$), the demand is concentrated in a set of sites and the population of these sites is divided into different strata depending on the kind of service that they require. The aim is to locate p centers to cover the different types of services demanded minimizing the weighted average of the largest distances associated with each of the different strata. In addition, it is considered that more than one stratum can be present at each site. Different formulations, valid inequalities and preprocessings are developed and compared for this problem. An application of this model is presented in order to implement a heuristic approach based on the Sample Average Approximation method (SAA) for solving the probabilistic p -center problem in an efficient way.

Keywords: Location, p -center, discrete optimization, Sample Average Approximation.

1 Introduction

Discrete location problems have been widely studied since the seminal paper Balinski (1965), where the first MILP formulation for such a problem was proposed. Among the fundamental problems in this area, the p -Center Problem (pCP) aims at selecting, from n given sites, the locations of p service centers that minimize the maximum distance between any of the sites and its closest service center. This model, in contraposition to the p -median problem, was motivated by the need not to discriminate spatially dispersed clients when locating essential or emergency centers (see Garfinkel et al., 1977; Calik et al., 2015, for more details).

Both, continuous and discrete versions have already been addressed by several authors. Examples of works concerning the continuous version are Callaghan et al. (2017) and Elshaikh et al. (2016) where the planar version is analyzed. This paper focuses on an extension of the discrete p CP. The discrete p CP, also known as vertex p CP, has been proven to be NP-hard (Kariv and Hakimi, 1979). However, many efficient exact and heuristic algorithms have been introduced for this problem. See for instance, Contardo et al. (2019); Calik and Tansel (2013) or Irawan et al. (2016).

In the last decades, several extensions of the discrete p CP have been introduced in the literature. These include variants considering capacities (Özsoy and Pınar, 2006; Albareda-Sambola et al., 2010; Quevedo-Orozco and Ríos-Mercado, 2015) or pre-existing centers, as in the conditional p -center problem (Drezner, 1989). Other extensions, such as the α -neighbor p -center problem, consider the largest distance of a demand point to its α closest centers, see Chen and Chen (2013).

In addition, the p CP with uncertain parameters has been addressed both, from the perspective of robust optimization (Averbakh and Berman, 1997; Lu and Sheu, 2013) and of stochastic programming (Revelle and Hogan, 1989; Espejo et al., 2015; Martínez-Merino et al., 2017). In particular, in Martínez-Merino et al. (2017) the probabilistic p -center problem (PpCP) is introduced. In this problem, the goal is to minimize the expected largest distance between any demand point and its corresponding center. The PpCP considers that the demand can occur independently at each demand site with a certain known probability. Observe, that the p CP is a particular case of this problem where all sites have demand with probability one. In this paper, we will introduce a heuristic method for the PpCP making use of the formulations of the p CP extension that will be proposed.

A common characteristic of most of the considered problem variants is that customers are assumed to be homogeneous in the sense that they are all considered in the same way in the objective function. The only exception would be the weighted p CP, where the distances between each site and its closest center are affected by site-dependent weights. See, for instance, Jeger and Kariv (1985) where the particular case of this problem defined on trees is addressed.

In this paper, we consider situations where, for instance, the population of a region is divided into different strata, and people of different strata can live together in the same

cities. The goal of the problem is to locate centers in such a way that the weighted sum of the largest distance associated with each stratum is minimized. This problem is called the Stratified p -Center Problem ($SpCP$) and it could be applied when the evaluation of the service is measured separately for each stratum due to social or political reasons. The idea of analyzing demands distributed in a spatially different way has been used in covering problems (Schilling et al., 1979) but, up to the best of our knowledge, it has not been applied in the context of the pCP .

A possible real life application could be the location of centers in an humanitarian relief planning framework. The purpose is to locate centers that provide different essential services and where not all demand points need all the services. This is the case of underdeveloped countries where humanitarian aid centers offering assistance (medical supplies, basic goods, clothes, etc.) need to be located. Note that in this context, the opening of many aid centers could be very costly, for instance, due to safety reasons. The model that we propose takes into account the largest distance associated with each of the provided services in contrast with the pCP , where only the largest distance is considered. Consequently, the $SpCP$ evaluates the performance of each of the services.

Another application of this model could be the location of warehouses for different perishable items (fruits, vegetables, seafood) whose demand sites are not the same for all the items. In this case, the warehouses should be located in such a way that the clients of each kind of item could be served as soon as possible to avoid the damage of the products.

Besides, this model could also be useful to locate social service centers that offer support to different social minorities needing assistance. The $SpCP$ model allows to minimize the weighted sum of the largest distance associated with each minority. In all the above mentioned applications, the weight associated with each stratum could be related to its importance or its associated cost.

The paper is organized as follows. In Section 2, a formulation for the $SpCP$ based on the Daskin (1995) and Calik and Tansel (2013) formulations for the pCP is introduced. In Section 3, some alternative formulations together with some valid inequalities are proposed. Section 4 applies the results in previous sections for an efficient implementation of a Sample Average Approximation heuristic for the probabilistic p -center problem (see Martínez-Merino et al., 2017). Section 5 reports the computational results comparing all the proposed formulations

and the results of using Sample Average Approximation. Finally, Section 6 gathers the conclusions of the paper.

2 Notation and classical formulation

Let $N = \{1, \dots, n\}$ be a given set of sites and $p \geq 2$ the number of facilities to be located. For each pair $i, j \in N$, let d_{ij} be the distance from location i to j . Besides, $d_{ii} = 0$ for $i \in N$ and $d_{ij} > 0$ for $i \neq j$. In the following we use the next notation. The sorted distances associated with pairs of sites are denoted by

$$0 = d_{(1)} < d_{(2)} < \dots < d_{(G)}.$$

The sorted distances from a site $i \in N$ to the remaining sites are denoted by

$$0 = d_{i(1)} < d_{i(2)} < \dots < d_{i(G_i)}.$$

In the previous notation G and G_i are the number of different distances between pairs of sites and between i and any other site, respectively (removing possible multiplicities).

Moreover, the population of each site $1, \dots, n$ is partitioned into a set of strata, taking into account that more than one stratum can be present in a site i and not always all the strata are present in a site. Given \mathcal{S} the set of strata in which the population is divided, we consider a family of subsets $\{N^s\}_{s \in \mathcal{S}}$ such that $N^s \subseteq N$ is the set of sites where stratum s is present for $s \in \mathcal{S}$. Then, the sorted distances from a stratum, i.e., the sorted sequence of family $\{d_{ij}\}_{i \in N^s, j \in N}$ is denoted by

$$0 = d_{(1)}^s < d_{(2)}^s < \dots < d_{(G^s)}^s,$$

where G^s is the number of different distances of the family $\{d_{ij}\}_{i \in N^s, j \in N}$.

The problem addressed in this work is based on the classical p CP. However, in contrast with the p CP, this new problem considers that population of the sites is divided in different strata depending on the kind of service that they require. For a given stratum s , this problem takes into account the largest distance from the sites where stratum s is present and their corresponding closest service facility. Recall that in the same site there can be inhabitants belonging to more than one stratum.

For each site $j \in N$, and each stratum $s \in \mathcal{S}$, the following binary parameter is defined:

$$\xi_j^s = \begin{cases} 1, & \text{if } j \in N^s, \\ 0, & \text{otherwise.} \end{cases}$$

Besides, each stratum has an associated weight, $(w_s, s \in \mathcal{S})$ that is used to balance the cost related to the different strata in the objective function. The weights can be interpreted in different ways. For instance, they can measure the importance given to a certain stratum.

Given the former parameters, the aim of this problem is to locate p service facilities minimizing the weighted sum of the largest assignments within each stratum. Therefore, the problem can be expressed synthetically in the following way:

$$\min_{\substack{P \subseteq N \\ |P|=p}} \sum_{s \in \mathcal{S}} w_s d(P, N^s), \quad (1)$$

where P is a subset of facilities to open and $d(P, N^s) = \max_{j \in N^s} \min_{i \in P} d_{ij}$. For a given site $j \in N$, we will refer to $\min_{i \in P} d_{ij}$ as the allocation distance of site j , so $d(P, N^s)$ is the maximum allocation distance among the sites with presence of stratum s , or equivalently within stratum s .

The problem previously described can be formulated using the classic p -center formulation (see Daskin, 1995). With this purpose, the following variables are defined:

$$x_{ij} = \begin{cases} 1, & \text{if site } j \text{ is assigned to center } i, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i, j \in N. \quad (2)$$

$$\theta^s = \text{largest allocation distance for the sites where stratum } s \text{ is present, } s \in \mathcal{S}. \quad (3)$$

Using these variables, the derived formulation is,

$$(F1) \quad \min \sum_{s \in \mathcal{S}} w_s \theta^s \quad (4)$$

$$\text{s.t.} \quad \sum_{i \in N} x_{ii} = p, \quad (5)$$

$$\sum_{i \in N} x_{ij} = 1, \quad j \in N, \quad (6)$$

$$x_{ij} \leq x_{ii}, \quad i, j \in N, \quad (7)$$

$$\theta^s \geq \sum_{i \in N} d_{ij} x_{ij}, \quad s \in \mathcal{S}, j \in N^s, \quad (8)$$

$$x_{ij} \in \{0, 1\}, \quad i, j \in N, \quad (9)$$

$$\theta^s \geq 0, \quad s \in \mathcal{S}. \quad (10)$$

Constraint (5) restricts that there are p centers. Constraints (6) indicate that each site is associated with only one center. Constraints (7) restrict that sites must be assigned to an open center. Constraints (8) ensure that the largest allocation distance within stratum s is not smaller than the allocation distance of any site where stratum s is present. As mentioned before, the objective function is the weighted sum of the largest distances within each stratum.

To the best of our knowledge, the most recent formulation for the p -center problem was given by Calik and Tansel (2013) providing very good results. We propose a formulation of our problem inspired in Calik and Tansel (2013) using the following families of variables.

$$\bar{u}_{sr} = \begin{cases} 1, & \text{if } d_{(r)} \text{ is the largest allocation distance among the sites in } N^s, \\ 0, & \text{otherwise,} \end{cases}$$

$$s \in \mathcal{S}, r = 1, \dots, G.$$

$$y_i = \begin{cases} 1, & \text{if a center is placed at } i, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i \in N.$$

Using these families of variables, the new formulation is given by

$$(F2) \quad \min \sum_{s \in \mathcal{S}} \sum_{k=1}^G w_s d_{(k)} \bar{u}_{sk} \quad (11)$$

$$\text{s.t.} \quad \sum_{i \in N} y_i = p, \quad (12)$$

$$\sum_{k=1}^G \bar{u}_{sk} = 1, \quad s \in \mathcal{S}, \quad (13)$$

$$\sum_{k'=1}^{k-1} \bar{u}_{sk'} \leq \sum_{\substack{i \in N \\ d_{ij} < d_{(k)}}} y_i, \quad s \in \mathcal{S}, j \in N^s, k = 2, \dots, G, \quad (14)$$

$$y_i \in \{0, 1\}, \quad i \in N, \quad (15)$$

$$\bar{u}_{sk} \in \{0, 1\}, \quad s \in \mathcal{S}, k = 1, \dots, G. \quad (16)$$

Constraint (12) restricts that there are p centers. Constraints (13) ensure that for each stratum, only one of the distances is the largest allocation distance. Constraints (14) determine that the largest allocation distance within a stratum s will be among the first k distances if there is a center with a distance smaller than or equal to $d_{(k)}$ with respect to any site in N^s .

Observe that \bar{u} -variables determine the largest allocation distance among the sites where each stratum $s \in \mathcal{S}$ is present. As a consequence, only the distances associated with sites in N^s will be necessary to obtain the largest distance with respect to s . Therefore, the number of variables can be reduced defining \tilde{u} -variables in the following way,

$$\tilde{u}_{sk} = \begin{cases} 1, & \text{if } d_{(k)}^s \text{ is the largest allocation distance for the sites in } N^s \\ 0, & \text{otherwise,} \end{cases} \quad s \in \mathcal{S}, k = 1, \dots, G^s.$$

Observe that in the original formulation F2, the number of \bar{u} -variables is $|\mathcal{S}|G$. However, by doing this reduction, the obtained number of variables is $\sum_{s \in \mathcal{S}} G^s$. Taking advantage of this reduction of the number of variables, the new objective function for the model is

$$\sum_{s \in \mathcal{S}} \sum_{k=1}^{G^s} w_s d_{(k)}^s \tilde{u}_{sk}, \quad (17)$$

and constraints (14) can be replaced by

$$\sum_{k'=1}^{k-1} \tilde{u}_{sk'} \leq \sum_{\substack{i \in N \\ d_{ij} < d_{(k)}^s}} y_i, \quad s \in \mathcal{S}, j \in N^s, k = 2, \dots, G^s. \quad (18)$$

Therefore, this new family of \tilde{u} -variables allows us to provide a new formulation with a smaller number of variables and constraints. Moreover, the following result allows to strengthen this new formulation.

Proposition 2.1 *For $s \in \mathcal{S}$ and $j \in N^s$, let $l_{jr}^s \in \{1, \dots, G^s\}$ be such that $d_{j(r)} = d_{(l_{jr}^s)}^s$. Considering formulation F2 with \tilde{u} variables (instead of \bar{u} variables), the objective function (17) and replacing (14) by*

$$\sum_{k'=1}^{l_{jr}^s-1} \tilde{u}_{sk'} \leq \sum_{\substack{i \in N \\ d_{ij} < d_{(l_{jr}^s)}^s}} y_i, \quad s \in \mathcal{S}, j \in N^s, r = 2, \dots, G_j, \quad (19)$$

results in a valid equivalent formulation F2' with a smaller number of constraints.

Proof:

We prove that constraint families (18) and (19) are equivalent. Let $\tilde{s} \in \mathcal{S}, \tilde{j} \in N^s$ and $\tilde{r} \in \{2, \dots, G_{\tilde{j}}\}$. Consider the following subset of constraints of family (18),

$$\sum_{k'=1}^{k-1} \tilde{u}_{\tilde{s}k'} \leq \sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(k)}^{\tilde{s}}}} y_i, \quad k \in \{l_{\tilde{j}, \tilde{r}-1}^{\tilde{s}} + 1, \dots, l_{\tilde{j}\tilde{r}}^{\tilde{s}}\}. \quad (20)$$

Observe that $\sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(l_{\tilde{j}, \tilde{r}-1}^{\tilde{s}} + 1)}^{\tilde{s}}}} y_i = \dots = \sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(l_{\tilde{j}\tilde{r}}^{\tilde{s}})}^{\tilde{s}}}} y_i$, then since

$$\sum_{k'=1}^{l_{\tilde{j}\tilde{r}-1}^{\tilde{s}}} \tilde{u}_{\tilde{s}k'} \leq \dots \leq \sum_{k'=1}^{l_{\tilde{j}\tilde{r}}^{\tilde{s}}-1} \tilde{u}_{\tilde{s}k'},$$

the family of constraints (20) is dominated by $\sum_{k'=1}^{l_{\tilde{j}\tilde{r}}^{\tilde{s}}-1} \tilde{u}_{\tilde{s}k'} \leq \sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(l_{\tilde{j}\tilde{r}}^{\tilde{s}})}^{\tilde{s}}}} y_i$.

Therefore, the obtained formulation F2' is equivalent to F2 with less constraints. In fact, the number of constraints (18) is $\sum_{j \in N} \sum_{s \in \mathcal{S}} \xi_j^s G^s$ and the number of constraints (19) is $\sum_{j \in N} \sum_{s \in \mathcal{S}} \xi_j^s G_j$. It is straightforward that for each pair, $s \in \mathcal{S}, j \in N^s$, $G_j \leq G^s$ since, at least, the distances associated with location j must be among the distances related to stratum s .

□

3 Formulation using covering variables

In this section we introduce three formulations making use of stratum-covering and site-covering variables. The idea behind these formulations is to take advantage of the information provided by considering the ordered sequence of possible assignment distances. In particular, the variables defined for these formulations determine whether the largest assignment distance associated with a stratum s is at least the one in a certain position of the sorted vector d^s (stratum-covering variables) and whether the allocation distance associated with a site i is at least the one in a certain position of the sorted vector d_i (site-covering variables). In this section we will see that the use of these variables associated with sorted vectors allows to propose new efficient formulations.

3.1 Formulation with stratum-covering variables

In this subsection we present a formulation based on the use of y -variables described in the previous section and the following family of variables:

$$u_{sk} = \begin{cases} 1, & \text{if the largest allocation distance for the sites in } N^s \text{ is at least } d_{(k)}^s, \\ 0, & \text{otherwise,} \end{cases}$$

for $s \in \mathcal{S}$, $k = 2, \dots, G^s$.

Observe that we have used the same strategy as in the former section, so that for each $s \in \mathcal{S}$ the number of u variables will be equal to the number of different distances associated with s . The use of this type of variables for the classical p CP was introduced by Elloumi et al. (2004). Inspired in this idea, we provide the following formulation for the Sp CP.

$$(F3) \quad \min \sum_{s \in \mathcal{S}} w_s \left(\sum_{k=2}^{G^s} (d_{(k)}^s - d_{(k-1)}^s) u_{sk} \right) \quad (21)$$

$$\text{s.t.} \quad \sum_{i \in N} y_i = p, \quad (22)$$

$$u_{sk} \geq 1 - \sum_{\substack{i \in N \\ d_{ij} < d_{(k)}^s}} y_i, \quad s \in \mathcal{S}, j \in N^s, k = 2, \dots, G^s, \quad (23)$$

$$y_i \in \{0, 1\}, \quad i \in N, \quad (24)$$

$$u_{sk} \in \{0, 1\}, \quad s \in \mathcal{S}, k = 2, \dots, G^s. \quad (25)$$

As it can be seen in (21), the objective function for this formulation can be expressed using a telescopic sum. Constraint (22) ensures that there are p open centers. Constraints (23) determine that if there is not a center at a distance smaller than $d_{(k)}^s$ from a site $j \in N^s$, then $u_{sk} = 1$.

Proposition 3.1 *Replacing (23) in F3 by the following families of constraints*

$$u_{s,l_{jr}^s} \geq 1 - \sum_{\substack{i \in N \\ d_{ij} < d_{j(r)}^s}} y_i, \quad s \in \mathcal{S}, j \in N^s, r = 2, \dots, G_j, \quad (26)$$

$$u_{sk} \leq u_{s,k-1}, \quad s \in \mathcal{S}, k = 3, \dots, G^s, \quad (27)$$

results in an equivalent formulation, F3-(23)+(26)+(27).

Proof:

Let $(\tilde{j}, \tilde{s}) \in N \times \mathcal{S}$ such that $\xi_{\tilde{j}}^{\tilde{s}} = 1$ and $\tilde{r}, \tilde{r} + 1 \in \{2, \dots, G_{\tilde{j}}\}$. Consider the following subset of constraints of family (23),

$$u_{\tilde{s}k} \geq 1 - \sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(k)}^{\tilde{s}}}} y_i, \quad k \in \{l_{\tilde{j}, \tilde{r}-1}^{\tilde{s}} + 1, \dots, l_{\tilde{j}\tilde{r}}^{\tilde{s}}\}. \quad (28)$$

Observe that $\sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(l_{\tilde{j}, \tilde{r}-1}^{\tilde{s}}+1)}^{\tilde{s}}}} y_i = \dots = \sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(l_{\tilde{j}\tilde{r}}^{\tilde{s}})}^{\tilde{s}}}} y_i$, then using (27), the family (28) is dominated by

$$u_{s, l_{\tilde{j}\tilde{r}}^{\tilde{s}}} \geq 1 - \sum_{\substack{i \in N \\ d_{i\tilde{j}} < d_{(l_{\tilde{j}\tilde{r}}^{\tilde{s}})}^{\tilde{s}}}} y_i.$$

□

Remark 3.1 *Formulation F3-(23)+(26)+(27) has a smaller number of constraints than F3 if*

$$\sum_{s \in \mathcal{S}} \left(\sum_{j \in N} \xi_j^s (G^s - G_j) - G^s + 2 \right) \geq 0.$$

3.2 Formulation with site-covering variables

In this section we propose a new formulation for our problem using the following set of variables, inspired in the ones defined by García et al. (2011) for the p CP:

$$z_{ir} = \begin{cases} 1, & \text{if the allocation distance of site } i \text{ is at least } d_{i(r)}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i \in N, r = 2, \dots, G_i.$$

Based in this set of variables and θ^s -variables defined by (3), we propose the following formulation for our problem:

$$(F4) \quad \min \sum_{s \in \mathcal{S}} w_s \theta^s$$

$$\text{s.t.} \quad \sum_{i \in N} z_{i2} = n - p, \quad (29)$$

$$\sum_{\substack{i \in N \\ d_{ij} < d_{j(r)}}} (1 - z_{i2}) \geq 1 - z_{jr}, \quad j \in N, r = 3, \dots, G_j \quad (30)$$

$$\theta^s \geq d_{j(r)} z_{jr}, \quad s \in \mathcal{S}, j \in N^s, r = 2, \dots, G_j, \quad (31)$$

$$z_{jr} \in \{0, 1\}, \quad j \in N, r = 2, \dots, G_j, \quad (32)$$

$$\theta^s \geq 0, \quad s \in \mathcal{S}. \quad (33)$$

Constraint (29) indicates that there are p centers. Constraints (30) ensure that if $z_{jr} = 0$ then, there is at least one center at i with $d_{ij} < d_{j(r)}$, i.e., location j is served by a center at a distance smaller than $d_{j(r)}$. Finally, constraints (31) ensure that θ^s is the largest allocation distance for sites in N^s .

Proposition 3.2 *Formulation F4 is still valid after relaxing the integrality of variables z_{ir} for $i \in N, r = 3, \dots, G_i$.*

Proof:

Let $(\tilde{\theta}, \tilde{z})$ be an optimal solution of F4 relaxing z_{ir} for $i \in N, r = 3, \dots, G_i$. We distinguish between two cases.

If $\sum_{\substack{i \in N \\ d_{i,i_0} < d_{i_0 r_0}}} (1 - \tilde{z}_{i2}) = 0$ then $\tilde{z}_{i_0 r_0} \geq 1$ due to constraints (30). Therefore, $\tilde{z}_{i_0 r_0} = 1$.

If $\sum_{\substack{i \in N \\ d_{i,i_0} < d_{i_0 r_0}}} (1 - \tilde{z}_{i2}) \geq 1$, then constraints (30) reduce to $z_{i_0 r_0} \geq 0$. Since positive values of $\tilde{z}_{i_0 r_0}$ penalize the objective function due to constraints (31), then $\tilde{z}_{i_0 r_0} = 0$. \square

Preliminary computational results show that this relaxation does not improve computational times of formulation F4.

Proposition 3.3 *Replacing constraints (31) in F4 by*

$$\theta^s \geq \sum_{r=2}^{G_j} (d_{j(r)} - d_{j(r-1)}) z_{jr}, \quad s \in \mathcal{S}, j \in N^s, \quad (34)$$

results in a valid formulation F4-(31)+(34) for the problem with less constraints, that dominates F4.

Proof:

Let $s \in \mathcal{S}, j \in N^s$. Note that, due to constraints (30) and constraints (34) it holds that $z_{jr} \leq z_{j,r-1}$ for $r \in \{3, \dots, G_j\}$ since,

$$\sum_{\substack{i \in N \\ d_{ij} < d_{j(r)}}} (1 - z_{i2}) \geq \sum_{\substack{i \in N \\ d_{ij} < d_{j(r-1)}}} (1 - z_{i2}),$$

and z -variables penalize in the objective function through constraints (34). Hence, since $z_{jr} \in \{0, 1\}$ we have that

$$\theta^s = \max_{j \in N^s} \left\{ \sum_{r=2}^{G_j} (d_{j(r)} - d_{j(r-1)}) z_{jr} \right\}$$

and then the formulation F4-(31)+(34) is valid. Moreover, for the relaxed problem we have that

$$\sum_{r=1}^{G_j} (d_{j(r)} - d_{j(r-1)}) z_{jr} \geq \max_{r=1, \dots, G_j} d_{j(r)} z_{jr}, \quad \forall s \in \mathcal{S}, j \in N^s,$$

i.e., this formulation dominates F4. Besides, the number of constraints (31) is $\sum_{j \in N} \sum_{s \in \mathcal{S}} \xi_j^s G_j$ and the number of constraints (34) is $\sum_{j \in N} \sum_{s \in \mathcal{S}} \xi_j^s$. Then, formulation F4-(31)+(34) has a smaller number of constraints than F4. \square

We have also studied alternative formulations using a non-cumulative version of the z -variables, i.e., defining

$$\bar{z}_{ir} = \begin{cases} 1, & \text{if the allocation distance of site } i \text{ is } d_{i(r)}, \\ 0, & \text{otherwise.} \end{cases} \quad \text{for } i \in N, r = 2, \dots, G_i.$$

Nevertheless, a preliminary computational analysis of these formulations shows a worse performance with respect to F4.

3.3 Formulation with stratum- and site-covering variables

The last formulation that we propose combines two families of covering variables, one associated with the distances from each stratum $s \in \mathcal{S}$ (u -variables) and another one with the

allocation of each site $i \in N$ (z -variables). The combination of both families of variables is inspired in the formulation of Marín et al. (2009) for the Discrete Ordered Median problem.

For each $s \in \mathcal{S}$, $k \in \{2, \dots, G^s\}$ and $i \in N$ we define

$$\bar{l}_{ik}^s = \begin{cases} r, & \text{if } r \in \{1, \dots, G_i\} \text{ exists such that } d_{i(r)} = d_{(k)}^s \text{ and } \xi_i^s = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the obtained formulation is

$$(F5) \quad \min \sum_{s \in \mathcal{S}} \sum_{k=2}^{G^s} w_s (d_{(k)}^s - d_{(k-1)}^s) u_{sk} \quad (35)$$

s.t. (29), (30),

$$u_{sk} \geq z_{i, \bar{l}_{ik}^s}, \quad s \in \mathcal{S}, i \in N^s, k = 2, \dots, G^s : \bar{l}_{ik}^s > 0, \quad (36)$$

$$u_{s, k-1} \geq u_{sk}, \quad s \in \mathcal{S}, k = 3, \dots, G^s, \quad (37)$$

$$u_{sk} \in \{0, 1\}, \quad s \in \mathcal{S}, k = 2, \dots, G^s, \quad (38)$$

$$z_{ir} \in \{0, 1\}, \quad i \in N, r = 2, \dots, G_i. \quad (39)$$

Constraints (36) determine the largest allocation distance among the sites in N^s . Observe that constraints (37) are valid inequalities for formulation F5. Indeed, if in a particular solution $u_{sk} = b$ and $u_{s, k-1} = a$ with $b > a$, then, a feasible solution with lower objective value can be found by taking $u_{sk} = a$. Constraints (37) are included in the formulation from the beginning since they provided good results in a preliminary computational study.

Note that constraints (36) can be equivalently written in the following way,

$$u_{s, l_{ir}^s} \geq z_{ir} \quad s \in \mathcal{S}, i \in N^s, r = 2, \dots, G_i. \quad (40)$$

Where l_{ir}^s is the index already defined in Proposition 2.1. To derive another valid formulation from (F5), we include the following notation,

$$l'_{ik}^s = \begin{cases} \min\{r : d_{i(r)} \geq d_{(k)}^s\}, & \text{if } d_{(k)}^s \leq d_{i(G_i)} \\ G_i + 1, & \text{otherwise.} \end{cases}$$

Proposition 3.4 *By replacing (36) in F5 by*

$$u_{sk} \geq z_{i, l'_{ik}^s}, \quad s \in \mathcal{S}, i \in N^s, k = 2, \dots, G^s, l'_{ik}^s \leq G_i. \quad (41)$$

a valid formulation, F5-(36)+(41), with a larger number of constraints is obtained.

Proof:

First, formulation F5-(36)+(41) is valid, since (41) determine the largest allocation distance among the sites where stratum s is present.

Observe that family of constraints (36) is a subset of constraints (41) since $l'_{ik} = \bar{l}_{ik}^s$ when $d_{i(r)} = d_{(k)}^s$ for some $r \in \{2, \dots, G_i\}$ and $\xi_i^s = 1$. Therefore F5-(36)+(41) dominates formulation F5. Concretely, the number of constraints (41) is $\sum_{i \in N} \sum_{s \in \mathcal{S}} \xi_i^s (G^s - 1)$. The number of constraints (36) is $\sum_{i \in N} \sum_{s \in \mathcal{S}} \xi_i^s (G_i - 1)$. As stated before $G_i \leq G^s$ for $s \in \mathcal{S}, i \in N^s$. Consequently, the number of constraints (41) is larger than the number of constraints (36). \square

Proposition 3.5 *i) Constraints (36) can be replaced by their following aggregated form:*

$$n_{sk} u_{sk} \geq \sum_{\substack{i \in N^s \\ \bar{l}_{ik}^s \neq 0}} z_{i\bar{l}_{ik}^s} \quad s \in \mathcal{S}, k = 2, \dots, G^s, \quad (42)$$

where $n_{sk} = |\{i \in N^s \text{ and there exists } r \in \{2, \dots, G_i\} \text{ such that } d_{i(r)} = d_{(k)}^s\}|$. This yields the new valid formulation, F5-(36)+(42).

ii) Constraints (41) can be replaced by their aggregated form that can be expressed as

$$n_s u_{sk} \geq \sum_{\substack{i \in N^s \\ l'_{ik} \leq G_i}} z_{i,l'_{ik}}, \quad s \in \mathcal{S}, k = 2, \dots, G^s, \quad (43)$$

Where $n_s = |N^s|$. This yields the new valid formulation F5-(36)+(43).

Proof:

i) Observe that, by (42), variables u_{sk} take the value 1 if the maximum distance among the sites in N^s is at least $d_{(k)}^s$. Indeed, if this allocation distance is at least $d_{(k)}^s$ then, by (30), there exists a site $j \in N^s$ such that $z_{j\bar{l}_{jk}^s} = 1$ and then, by (42), $u_{sk} = 1$.

Moreover, (42) are valid since n_{sk} is the maximum value that the right hand side of constraints (42) can take.

ii) By an argument analogous to the one discussed in i), we have that formulation F5-(36)+(43) is valid for the SpCP.

□

Besides, another aggregated version of constraints (36) is:

$$\sum_{\substack{s \in \mathcal{S}: \xi_i^s = 1 \\ l_{ir}^s \geq 2}} u_{s, l_{ir}^s} \geq \left(\sum_{s \in \mathcal{S}} \xi_i^s \right) z_{ir} \quad i \in N, r = 2, \dots, G_i. \quad (44)$$

Some computational studies have been carried out with formulation F5-(36)+(44). However, it provides worse running times than formulations presented in Proposition 3.5.

Proposition 3.6 *Formulation F5 and all its variants (F5-(36)+(40), F5-(36)+(41), F5-(36)+(42), F5-(36)+(43), F5-(36)+(44)) remain valid if integrality of variables z_{ir} is relaxed for $i \in N, r = 3, \dots, G_i$.*

Proof:

Let (\tilde{u}, \tilde{z}) be an optimal solution of the model relaxing z_{ir} for $i \in N, r = 3, \dots, G_i$. If (\tilde{u}, \tilde{z}) are all binary, we are done. Otherwise, there is at least one $0 < \tilde{z}_{i_0 r_0} < 1$ with $i_0 \in N, r_0 \in \{3, \dots, G_{i_0}\}$. For this variable, Constraint (30) reduces to $z_{i_0 r_0} \geq 0$, since z_{i_2} are binary for $i \in N$. Hence $\tilde{z}_{i_0 r_0}$ value can be replaced by 0 without violating these constraints. Besides, constraints (36), (40), (41), (42), (43) or (44) (depending on the variant of F5) are not violated if $\tilde{z}_{i_0 r_0}$ takes value 0 and the objective value is not worse. □

Computational results in Section 5 show that this relaxation improves the times of formulation F5-(36)+(43).

Proposition 3.7 *Formulations F5, F5-(36)+(40) and F5-(36)+(41), remain valid if we relax the integrality condition of u_{sk} variables for $s \in \mathcal{S}, k = 2, \dots, G^s$ and z_{ir} variables for $i \in N, r \in \{3, \dots, G_i\}$.*

Proof:

Since z -variables take integer values as observed in Proposition 3.6 and since u_{sk} for $s \in \mathcal{S}, k = 2, \dots, G^s$ penalize the objective function, it holds that u_{sk} take integer values due to constraints (36) (or equivalently, due to constraints (40) or (41)). □

Preliminary computational results show that the relaxations introduced in Proposition 3.7 do not improve the running times of the corresponding models.

3.4 Reducing the number of covering variables

Observe that some of z -variables described in formulations of subsections 3.2 and 3.3 could be fixed. Since p centers are located in the $SpCP$, then the distance associated with a client i will not be among the $p - 1$ worst possible ones. Then, the following constraints allow to fix some variables.

Let $\tilde{d}_{i(1)} \leq \tilde{d}_{i(2)} \leq \dots \leq \tilde{d}_{i\tilde{G}_i}$ be the sorted distances of all possible assignments of site i (observe that this sequence of distances can contain repeated values), then

$$z_{ir} = 0 \quad \forall i \in N, r \in \{2, \dots, G_i\} \text{ such that } d_{i(r)} > \tilde{d}_{i(n-p+1)}. \quad (45)$$

Consequently, for each $i \in N$ it is only necessary to define z_{ir} for $r = 2, \dots, G_i$ such that $d_{i(r)} \leq \tilde{d}_{i(n-p+1)}$.

Regarding u -variables appearing in formulations F3 and F5, observe that these are binary variables indicating for each stratum $s \in \mathcal{S}$ whether the largest distance associated with stratum s is at least $d_{(k)}^s$ or not, where $k = 2, \dots, G^s$. The number of u -variables for each stratum $s \in \mathcal{S}$ is $G^s - 1$, i.e, the number of different distances from sites in N^s to all candidate locations (excluding distance 0). In this subsection, we analyze if the number of u -variables can be reduced for each stratum.

In fact, the number of u -variables could be reduced if tighter bounds on the largest allocation distance associated with each stratum for the $SpCP$ were known. The following proposition exploits this argument.

Proposition 3.8 *For each stratum $s \in \mathcal{S}$, let $v(pCP_s)$ be the optimal value of a p -center problem where the set of candidates centers is N and the set of demand points is N^s , from now on, denoted with pCP_s . Then, the largest allocation distance associated with s is at least $v(pCP_s)$ in the optimal solution of the $SpCP$.*

Proof:

Observe that the solution of the $SpCP$ is feasible for the pCP_s . Then, given a solution of $SpCP$, its objective value for pCP_s will be greater than or equal to $v(pCP_s)$. \square

As a result, if a lower bound or the optimal value of pCP_s is obtained, then the number of u -variables associated with stratum s can be reduced. To reduce the number of variables we can follow the next scheme for each $s \in \mathcal{S}$:

- Obtain a lower bound on the pCP_s or its optimal objective value. This value can be denoted as LB_s .
- Define u_{sk} variables for all $k \in \{h : 2 \leq h \leq G^s \text{ and } d_{(h)}^s > LB_s\}$.
- For each $s \in \mathcal{S}$, given that $d_{(k_s)}^s$ is the largest distance associated with stratum s such that $d_{(k_s)}^s \leq LB_s$, the considered objective function will be:

$$\sum_{s \in \mathcal{S}} w_s \sum_{k=k_s+1}^{G^s} \left(d_{(k_s)}^s + (d_{(k)}^s - d_{(k-1)}^s) u_{sk} \right)$$

Observe that this is equivalent to fix $u_{sk} = 1$ for $k \leq k_s, s \in \mathcal{S}$.

Several criteria can be used to obtain an adequate bound LB_s for each stratum. In particular, in the computational experiments of this work we present two ways for obtaining these bounds. The first one uses the linear relaxation of the pCP_s using the classic formulation of Daskin (1995). The second one consists in using the binary algorithm proposed in Calik and Tansel (2013).

Observe that the argument described in constraints (45) for z -variables could be also useful to fix some of the u -variables. In particular, the following variables can be fixed:

$$u_{sk} = 0, (s, k) \in \mathcal{K}, \quad (46)$$

where \mathcal{K} is the set of pairs $(s, k) \in \mathcal{S} \times \{2, \dots, G^s\}$ such that for every $i \in N^s$, $d_{(k)}^s > \tilde{d}_{i(n-p+1)}$.

Summing up, u -variables can be reduced using the scheme described before and constraints (46). In section 5 we study the percentage of z - and u - variables fixed by applying the former criteria.

3.5 Valid inequalities for F5

Some constraints related to closest assignments could be applied for this problem. Some of the constraints appearing in Espejo et al. (2012) have been adapted for formulation F5 (the most promising formulation as we will see in Section 5). However, the only valid inequality that presents good results is the one described below:

$$z_{ir} \leq z_{j2} \quad i, j \in N, r = 2, \dots, G_i : d_{i(r-1)} = d_{ij}, \quad (47)$$

These constraints could be considered as derived from the ones proposed by Dobson and Karmarkar (1987). As observed, given $i, j \in N$ these constraints restrict the distance associated with i to be smaller than or equal to the distance d_{ij} if a center is located at j .

In the following we introduce other valid inequalities that take advantage of the relationship between two different strata.

$$\sum_{k=2}^{G^{s_1}} (d_{(k)}^{s_1} - d_{(k-1)}^{s_1}) u_{s_1 k} \leq \sum_{k=2}^{G^{s_2}} (d_{(k)}^{s_1} - d_{(k-1)}^{s_1}) u_{s_2 k}, \quad s_1, s_2 \in S : N^{s_1} \subseteq N^{s_2}. \quad (48)$$

These constraints state that the largest allocation distance associated with stratum s_1 will be smaller than or equal to the one associated with stratum s_2 if stratum s_2 is present in each site of N^{s_1} . Similarly the next constraints follow:

$$u_{s_1 k} \leq u_{s_2 l}, \quad s_1, s_2 \in S, k = 2, \dots, G^{s_1}, l = 2, \dots, G^{s_2} : N^{s_1} \subseteq N^{s_2}, d_{(k)}^{s_1} = d_{(l)}^{s_2}, \quad (49)$$

Constraints (49) hold since if the largest allocation distance associated with s_2 is smaller than $d_{(l)}^{s_2}$ and $N^{s_1} \subseteq N^{s_2}$, then the largest allocation distance within s_1 cannot be greater than or equal to $d_{(l)}^{s_2} = d_{(k)}^{s_1}$. The accumulated version of these valid inequalities is:

$$\sum_{k=2}^{G^{s_1}} u_{s_1 k} \leq \sum_{k=2}^{G^{s_2}} u_{s_2 k}, \quad s_1, s_2 \in S : N^{s_1} \subseteq N^{s_2}. \quad (50)$$

Other valid inequalities are those ensuring that z variables are sorted in non-increasing order for each $i \in N$, i.e.,

$$z_{ir} \geq z_{i,r+1}, \quad i \in N, r = 2, \dots, G_i - 1. \quad (51)$$

All these valid inequalities will be analyzed in Section 5.

4 Using SpCP to implement a SAA for solving the PpCP

Recall from Martínez-Merino et al. (2017) that the Probabilistic p CP (PpCP) is defined as the variant of the p CP where sites represent potential demand points, and the locations of the p centers have to be decided before the actual subset of sites that need to be served is revealed. In this problem, the goal is to minimize the expected maximum distance between a site with demand and its closest center. Here, expectation is computed with respect to the

probability distribution of the binary random vector defining the subset of sites that have demand.

Notice that, in fact, when uncertainty is modeled by means of a set of scenarios, the PpCP can be cast as a SpCP. In this case, each stratum would represent the set of sites having demand at a given scenario, and the stratum weight would correspond to the corresponding scenario probability. This suggests exploiting the SpCP formulations presented in this paper to solve the PpCP using the well-known Sample Average Approximation method (SAA).

SAA is based on using Monte Carlo Sampling in the probability space defined by the random variables involved in a problem definition (see Homem-de-Mello and Bayraksan, 2014). Although this idea was already used before for solving stochastic programming problems (Rubinstein and Shapiro, 1990; Robinson, 1996), the term SAA was formally defined in Kleywegt et al. (2002). We next provide a sketch of this methodology; for more details, see Shapiro (2013) or Linderoth et al. (2006).

Consider the two stage program (P) $z^* = \min_{x \in X} f(x) + Q(x)$, where the recourse function is defined as $Q(x) = \mathbb{E}_\xi[v(x, \xi)]$ and, given a solution x and a realization of the random vector ξ , ξ_0 , the so-called second stage problem is $v(x, \xi_0) = \min_{y \in Y(x, \xi_0)} q(y; x, \xi_0)$. Note that if ξ is a discrete random vector with a finite support, Ω , and each scenario $s \in \Omega$ has a known probability p^s , then, by replicating the variables of the second stage problem, (P) can be equivalently expressed as:

$$\begin{aligned} (P') \quad z^* = \min & f(x) + \sum_{s \in \Omega} p^s q(y^s; x, \xi^s) \\ \text{s.t.} \quad & x \in X, \\ & y^s \in Y(x, \xi^s), \quad s \in \Omega. \end{aligned} \tag{52}$$

Accordingly, using a random sample $\Omega^M \subset \Omega$, with $M = |\Omega^M|$, P can be approximated as

$$\begin{aligned} (P^M) \quad z^M = \min & f(x) + \frac{1}{M} \sum_{s \in \Omega^M} q(y^s; x, \xi^s) \\ \text{s.t.} \quad & x \in X, \\ & y^s \in Y(x, \xi^s), \quad s \in \Omega^M. \end{aligned} \tag{53}$$

Problem P^M is often referred to as sample average approximation problem. It is well known that given M , the expected value of this problem, $\mathbb{E}(z^M)$, is a lower bound on z^* and

it converges to z^* as N increases. Moreover, under some mild conditions on X and v , the random vector $x^{M,*}$ representing the optimal solution to P^M becomes arbitrarily close to the set of optimal solutions to P with probability 1. A common way to estimate $\mathbb{E}(z^M)$ is to solve a sequence of realizations of P^M for a given sample size M , and use the average of the corresponding optimal values as an estimate of $\mathbb{E}(z^M)$. The sequence is evaluated iteratively, and the termination criterion is most often related with the convergence of this average. The best of the solutions obtained in that sequence of problems is kept as a good approximation of the optimal solution.

Compared to other heuristics, the main advantage that SAA provides is the theoretical results that ensure the convergence of the method. In the case of the PpCP, we propose a classical SAA method in which the novelty is that the resulting problem in each iteration is a SpCP. Consequently, we can exploit the characteristics of the best formulations for the SpCP to enhance the performance of the SAA method. The pseudocode given in Algorithm 1 describes the SAA for case of the PpCP.

In the next section we will show some computational results of SAA using random samples of size $M = 10$. Besides, we will see how the use of different formulations of SpCP can affect the performance of the SAA.

5 Computational results

This section is devoted to the computational studies of the formulations described along the paper for the SpCP. The instances used in this computational experience are based on the p -median instances from the ORLIB¹.

For the smallest instances ($n = 6, \dots, 75$), the used matrices are submatrices of instances pmed1, pmed2, pmed3, pmed4 and pmed5 from the ORLIB data. For instances with $n = 100, 200, 300, 400$, the matrices are those corresponding to instances pmed1-pmed20. In all cases, several p values are considered ranging between $p = 2$ (for the smallest instances) to $p = 60$ (for the largest instances). Finally, in Table 7 all the ORLIB distance matrices together with their corresponding p values are studied.

¹Electronically available at <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/>

Algorithm 1: SAA for the PpCP.

```

/*  $K$  is a maximum number of iterations and,  $Av_{last}$  and  $Av_{new}$  are the
   average of the objective value solutions related to the last
   iterations (conveniently initialized to 1 and 2, respectively). */
 $k := 1$ ,  $Av_{last} := 1$ ,  $Av_{new} := 2$ ,  $K := 500$ .
while  $k < K$  and  $|Av_{new} - Av_{last}| > 0.0005 \cdot Av_{last}$  do
    /* Step 1: Generation of a random sample  $\Omega_k \in \Omega$ . */
    for  $k' = 1$  to  $k' = M$  do
        for  $i = 1$  to  $i = n$  do
            Create a random number  $r \in [0, 1)$ . Let  $q_i$  be the probability of client  $i$  to
            have demand.
            if  $r < q_i$  then
                 $\xi_i^{k'} := 1$ ,
            end
            else
                 $\xi_i^{k'} := 0$ .
            end
             $s^{k'} := (\xi_1^{k'}, \xi_2^{k'}, \dots, \xi_n^{k'})$ 
        end
    end
     $\Omega^k = \{s^1, \dots, s^M\}$ .
    /* Step 2: Solving of the sample average approximation problem. */
    Solve the SpCP where  $\mathcal{S} = \Omega^k$  and  $w_{s^{k'}} = \frac{1}{M}$  for  $k' = 1, \dots, M$ . Use one of the
    formulations in sections 2 or 3. Denote by  $val$  the optimal objective value of this
    problem.
    /* Step 3: Evaluation of the solution. */
    /*  $Av_{last}$  and  $Av_{new}$  allow to compare the average of objective values
       after a number of iterations: */
     $Av_{last} :=$  Average of the optimal objective values in the last  $k - 1$  iterations.
     $Av_{new} :=$  Average of the optimal objective values in the last  $k$  iterations.
    Fix the solution of the SpCP in the objective function of the PpCP obtaining an
    upper bound for the PpCP. ( $UB$ )
    if  $k = 1$  then
        |  $UB_{best} := UB$ 
    end
    else
        if  $UB < UB_{best}$  then
            |  $UB_{best} := UB$ .
        end
    end
end
end

```

For each instance, a total of $|\mathcal{S}| = 10$ strata are generated. Besides, each stratum (s) is independently created. First, a number $q_i \in (0, 1)$ is associated with each $i \in N$. Then a random number in $r \in [0, 1)$ is created. If $r < q_i$, then $\xi_i^s = 1$. Otherwise, $\xi_i^s = 0$.

The formulations are implemented in the commercial solver Xpress 8.0 using the modeling language Mosel. All the runs are carried out on the same computer with an Intel(R) Core(TM) i7-4790K processor with 32 GB RAM. We remark that the cut generation of Xpress is disabled to compare the relative performance of formulations cleanly.

First, we report a comparison of all proposed formulations in sections 2 and 3. In this study, we observe that the best results are provided by a variant of formulation F5. After that, we analyze if valid inequalities and the reduction of variables improve the computational times. Finally, Sample Average Approximation for PpCP is implemented using some of the SpCP formulations presented before.

5.1 Comparison of formulations

Before the comparison of the different formulations, we include an example along with its data that illustrates a solution of the SpCP for a specific instance.

Example 5.1 *Let N be a set of sites with $|N| = 10$ in which the distances between each pair of sites are given by the next distance matrix:*

$$d = \begin{pmatrix} 0 & 77 & 139 & 135 & 157 & 174 & 193 & 204 & 206 & 209 \\ 77 & 0 & 62 & 107 & 129 & 146 & 161 & 150 & 146 & 149 \\ 139 & 62 & 0 & 90 & 112 & 129 & 117 & 106 & 102 & 105 \\ 135 & 107 & 90 & 0 & 22 & 39 & 58 & 69 & 73 & 76 \\ 157 & 129 & 112 & 22 & 0 & 17 & 36 & 47 & 51 & 54 \\ 174 & 146 & 129 & 39 & 17 & 0 & 19 & 30 & 34 & 37 \\ 193 & 161 & 117 & 58 & 36 & 19 & 0 & 11 & 15 & 18 \\ 204 & 150 & 106 & 69 & 47 & 30 & 11 & 0 & 4 & 7 \\ 206 & 146 & 102 & 73 & 51 & 34 & 15 & 4 & 0 & 3 \\ 209 & 149 & 105 & 76 & 54 & 37 & 18 & 7 & 3 & 0 \end{pmatrix}.$$

Besides, the demand is divided into ten strata. Each stratum is present in a subset of sites as shown in Table 1. This table also includes the weight associated with each stratum.

Table 1: Strata data for Example 5.1.

sites	weight		sites	weight
s_1 : {1,2,3,6,7,9}	w_1 : 0.05		s_6 : {2,3,6,8,10}	w_6 : 0.05
s_2 : {6,8}	w_2 : 0.1		s_7 : {3,4,5,6,8,9,10}	w_7 : 0.05
s_3 : {6,9}	w_3 : 0.1		s_8 : {2,5,7}	w_8 : 0.05
s_4 : {6,7}	w_4 : 0.1		s_9 : {5,7,8,10}	w_9 : 0.1
s_5 : {2,5,10}	w_5 : 0.3		s_{10} : {2,6,7,8}	w_{10} : 0.1

The solution for this example is to open the facilities at sites 2, 5 and 10 with an objective value of 19.75. Consequently, demand points 1, 2 and 3 will be served by the center in 2, points 4, 5 and 6 will be covered by the facility in 5 and the remaining sites will be served by facility opened in 10. The largest distances associated with each stratum will be the ones given in the first row of Table 2.

Table 2: Largest distance associated with each stratum for the optimal solution (first row) and another feasible solution (second row) in Example 5.1.

Solution\ stratum	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
{2,5,10}	77	17	17	18	0	62	62	18	18	18
{1,2,6}	62	30	34	19	37	62	62	19	37	30

Observe that this solution for the Stratified p -center problem is not an optimal solution for the p -center problem. Particularly, the optimal solution for the p -center is the location of centers in sites 1, 2 and 6. The second row of Table 2 shows the largest distance associated with each stratum when using the optimal solution of the p -center problem. Note that the maximum distances in the strata with larger weights (s_2 , s_3 , s_4 , s_5 , s_9 and s_{10}) are reduced if the Stratified p -center solution is used. We can conclude that SpCP is worth it in order to obtain a better average performance among the strata.

Table 3 reports the results of the SpCP formulations proposed in sections 2 and 3. As can be observed, some formulations include several variants replacing some of the constraints by others. With these new constraints, the aim is to improve the running times of some of

these formulations. Table 3 reports two columns for each formulation. The first one shows the average running time for solving the model and the second column reports the average LP gap (in percentage, %). The LP gap is calculated as $\frac{OPT - LP}{OPT} \cdot 100$, where OPT is the optimal objective value and LP is the objective value of its linear relaxation. Observe that OPT is known for all the instances used in Table 3 since F5-(36)+(43)* model allows to solve them in less than two hours.

Note that each entry corresponds to the average over five instances of the same size and that the reported average running time is the average among the instances that are solved in less than two hours. The number of unsolved instances after two hours is reported in parentheses. In the LP gap column, the average final gaps for those instances that were not solved in two hours is shown in parentheses. This final gap is obtained as $\frac{UB_{best} - LB_{best}}{UB_{best}} \cdot 100$ where UB_{best} is the objective value of the best feasible solution obtained in two hours and LB_{best} is the best lower bound obtained in two hours. Besides, observe that formulation F5-(36)+(43)* corresponds to formulation F5 replacing constraints (36) by constraints (43) relaxing variables z_{ir} for $i \in N$, $r \in \{3, \dots, G_i\}$.

In terms of running times, observe that for $n = 100$ some of the instances cannot be solved in less than two hours if formulations F1, F2 or F3 are used. However, the reported results of F2' and F3-(23)+(26)+(27) are much better than those corresponding to F2 or F3. Note that times of F4 are similar in many of the cases to those required by F3-(23)+(26)+(27) and all the instances can be solved in less than two hours.

Observe also that F5 seems to provide better results than F4. Furthermore, it is clear that the best formulation is F5 replacing constraints (36) by constraints (43) and relaxing the integrality of variables z_{ir} for $i \in N$, $r \in \{3, \dots, G_i\}$. By using this variant of formulation F5, the results show that running times are (in average) not bigger than 65 seconds in any of the cases.

In contrast, the LP gaps of F2, F2', F3, F3-(23)+(26)+(27), F5 and F5-(36)+(41), which always coincide, are the smallest ones. Although F5-(36)+(43)* is the formulation that provides the best computational times, the reported LP gaps are the largest ones if we compare them with the remaining formulations.

Since F5-(36)+(43)* is the best formulation in terms of times, next subsection is devoted to the computational study of this formulation reducing the number of variables and using

Table 3: Formulations times and LP gap comparison

n	p	F1		F2		F2'		F3		F3-(23)+(36)+(27)		F4		F5		F5-(36)+(41)		F5-(36)+(42)		F5-(36)+(43)		F5-(36)+(43)*	
		Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap	Time	LP Gap
6	2	0.00	20.96	0.02	11.12	0.01	11.12	0.02	11.12	0.02	11.12	0.01	33.77	0.02	11.12	0.02	11.12	0.01	24.20	0.01	50.73	0.01	50.73
10	3	0.02	28.50	0.16	19.20	0.05	19.20	0.19	19.20	0.08	19.20	0.03	40.80	0.08	19.20	0.16	19.20	0.09	26.82	0.05	58.92	0.06	58.92
10	5	0.01	44.34	0.13	28.72	0.04	28.72	0.15	28.72	0.05	28.72	0.03	52.08	0.05	28.72	0.09	28.72	0.05	43.40	0.05	74.60	0.05	74.60
13	3	0.06	30.20	0.89	17.60	0.18	17.60	0.61	17.60	0.23	17.60	0.09	47.82	0.23	17.60	0.56	17.60	0.26	29.86	0.12	58.12	0.11	58.12
13	5	0.04	40.94	0.55	22.38	0.10	22.38	0.37	22.38	0.12	22.38	0.07	49.22	0.12	22.38	0.21	22.38	0.13	35.13	0.09	65.30	0.09	65.30
13	8	0.03	48.74	0.28	24.11	0.05	24.11	0.19	24.11	0.07	24.11	0.05	51.34	0.06	24.11	0.11	24.11	0.06	39.06	0.05	78.82	0.05	78.82
15	3	0.10	30.14	1.39	16.08	0.31	16.08	1.40	16.08	0.47	16.08	0.14	48.61	0.38	16.08	0.75	16.08	0.40	27.60	0.17	56.42	0.18	56.42
15	7	0.06	53.11	0.93	32.47	0.15	32.47	0.68	32.47	0.21	32.47	0.09	58.53	0.16	32.47	0.35	32.47	0.18	42.18	0.14	75.30	0.14	75.30
15	10	0.04	46.31	0.47	19.74	0.06	19.74	0.29	19.74	0.08	19.74	0.07	49.00	0.06	19.74	0.14	19.74	0.05	38.90	0.07	83.32	0.06	83.32
20	3	0.25	33.59	6.92	18.87	0.94	18.87	5.12	18.87	1.27	18.87	0.31	51.65	0.85	18.87	2.43	18.87	1.32	28.76	0.47	54.20	0.51	54.20
20	7	0.21	44.51	3.35	21.66	0.43	21.66	1.80	21.66	0.56	21.66	0.25	52.90	0.38	21.66	0.94	21.66	0.56	32.66	0.29	68.11	0.31	68.11
20	10	0.14	52.88	2.43	27.43	0.32	27.43	1.88	27.43	0.43	27.43	0.20	58.24	0.28	27.43	0.56	27.43	0.37	40.04	0.29	77.68	0.29	77.68
25	3	0.54	29.89	17.66	15.68	1.88	15.68	11.86	15.68	2.19	15.68	0.57	50.16	1.23	15.68	4.06	15.68	2.34	26.84	1.02	50.37	1.13	50.37
25	7	0.56	43.51	15.33	22.04	1.36	22.04	11.27	22.04	1.84	22.04	0.52	53.50	0.95	22.04	3.10	22.04	1.55	32.93	0.65	63.56	0.73	63.56
25	10	0.53	55.26	12.37	30.77	1.49	30.77	13.65	30.77	1.76	30.77	0.46	61.48	1.13	30.77	2.58	30.77	1.39	43.34	0.80	74.92	0.79	74.92
30	3	1.03	27.94	47.60	13.29	3.93	13.29	36.80	13.29	4.89	13.29	1.04	48.61	2.15	13.29	6.94	13.29	3.94	28.05	1.66	52.20	1.55	52.20
30	7	1.17	39.24	25.19	18.52	2.37	18.52	27.59	18.52	2.80	18.52	0.83	51.50	1.70	18.52	5.24	18.52	2.28	35.45	0.97	63.02	0.95	63.02
30	10	1.14	50.88	34.23	26.21	3.05	26.21	42.51	26.21	3.93	26.21	0.95	58.81	1.67	26.21	5.11	26.21	2.35	42.75	1.14	70.99	1.24	70.99
40	3	3.12	26.89	148.29	12.97	13.13	12.97	152.82	12.97	7.88	12.97	3.13	48.55	4.85	12.97	11.46	12.97	7.69	31.65	3.25	51.45	2.75	51.45
40	7	7.41	37.54	146.75	16.74	11.45	16.74	132.06	16.74	12.42	16.74	2.83	51.51	4.65	16.74	20.34	16.74	8.56	34.90	2.62	58.41	2.77	58.41
40	10	5.85	44.38	123.78	21.11	14.14	21.11	179.14	21.11	11.45	21.11	2.91	55.03	4.09	21.11	17.93	21.11	7.09	37.62	2.28	65.49	2.54	65.49
50	5	11.72	30.44	314.63	12.63	38.32	12.63	385.67	12.63	19.99	12.63	9.10	49.44	7.86	12.63	34.72	12.63	15.52	38.65	5.46	54.26	5.50	54.26
50	10	45.50	42.04	565.37	20.25	60.62	20.25	1357.87	20.25	43.47	20.25	18.39	54.50	12.96	20.25	70.69	20.25	17.71	43.19	7.79	62.29	5.94	62.29
50	15	57.11	47.35	294.64	20.88	32.18	20.88	615.38	20.88	41.99	20.88	9.50	57.07	10.91	20.88	36.25	20.88	15.87	44.26	6.54	69.01	8.80	69.01
75	5	110.47	28.03	1779.03	11.50	292.01	11.50	1304.42(1)	11.50 (1.3)	131.39	11.50	165.08	48.94	47.01	11.50	442.62	11.50	52.17	48.03	24.48	18.73	54.79	54.79
75	10	927.58	36.27	2518.22	13.98	416.00	13.98	3519.38(2)	13.98 (9.5)	173.89	13.98	239.72	51.55	100.70	13.98	239.59	13.98	66.51	50.36	26.83	19.97	59.35	59.35
75	15	1973.03	41.65	1931.90	16.57	380.97	16.57	3567.20(2)	16.57 (8.8)	231.85	16.57	101.01	53.75	59.88	16.57	201.43	16.57	73.25	54.31	28.21	23.56	64.43	64.43
100	10	5981.32(4)	26.05 (2.7)	3443.88(1)	12.15 (7.8)	1090.56	12.15	(5)	12.15 (10.7)	476.64	12.15	457.77	50.15	164.32	12.15	605.83	12.15	186.07	56.74	63.53	59.06	64.09	59.06
100	15	(5)	37.41 (5.4)	2260.86(2)	13.17 (9.2)	1611.93	13.17	3277.27(4)	13.17 (13.1)	476.51	13.17	958.80	52.14	142.57	13.17	435.76	13.17	271.07	57.95	77.88	64.12	61.87	61.87
100	25	5098.56(4)	49.16 (14)	4780.63(2)	19.07 (14.7)	1212.76	19.07	2887.83(4)	19.07 (13.3)	968.22	19.07	544.08	58.96	133.52	19.07	674.77	19.07	336.73	61.62	78.34	64.62	69.13	69.13

valid inequalities.

5.2 Reduction of variables and valid inequalities for F5-(36)+(43)*

In this subsection we observe the results of using a preprocessing phase to reduce the number of variables in formulation F5-(36)+(43)* and we will also report the results when applying valid inequalities.

In Subsection 3.4 a preprocessing phase to reduce the number of z - and u -variables is described. Concretely, constraints (45) allow to reduce the number of defined z -variables. Similarly, constraints (46) decrease the number of u -variables. Besides, a reduction of u -variables based on obtaining an adequate lower bound of the p -center objective value considering each stratum independently is described.

Table 4: Percentage of z - and u -variables reduced with respect to the original ones.

		% z	% u	
n	p	(45)	(46)+clas. Rel	(46)+Binary alg.
75	5	6.35	29.47	43.11
75	10	13.61	23.58	36.63
75	15	20.21	21.75	32.79
100	10	10.74	25.44	37.32
100	15	16.13	23.75	34.84
100	25	26.01	22.55	31.68
200	10	9.26	30.23	40.72
200	20	16.93	25.66	35.33
200	30	22.88	24.82	33.96

In particular, we mention two ways to obtain these lower bounds. The first one is to solve the linear relaxation for the pCP using the classic formulation of Daskin (1995). The second way consists in using the binary algorithm proposed in Calik and Tansel (2013). Table 4 reports the percentage of fixed z - and u -variables in formulation F5-(36)+(43)* when the former criteria for fixing variables are applied. The first column corresponds to the percentage of reduced z -variables if constraints (45) are applied. The second column reports the percentage of fixed u -variables when using constraints (46) together with the reduction strategy based on the solving of Daskin (1995) relaxed formulation for each stratum. Finally

the last column reports the percentage of reduction when (46) and Binary Algorithm specified in Calik and Tansel (2013) for each stratum are applied. Observe that between 6.35% and 26.01% of the z -variables could be fixed. In the case of u -variables the largest number of fixed u -variables (boldfaced) is obtained when applying the Binary Algorithm. With this strategy and (46), more than a 31% of u -variables are fixed in average.

Table 5: Times and LP gaps reducing the number of z - and u -variables in formulation F5-(36)+(43).

n	p	F5-(36)+(43)		F5-(36)+(43)*		Classic rel			Binary			Binary*		
		Time	LP Gap	Time	LP Gap	t prepro	t total	LP Gap	t prepro	t total	LP Gap	t prepro	t total	LP Gap
75	5	24.48	54.79	18.73	54.79	1.53	19.53	33.56	0.52	12.50	8.55	0.51	7.95	8.55
75	10	26.83	59.35	19.97	59.35	1.09	16.60	41.45	0.49	13.56	10.82	0.49	11.04	10.82
75	15	28.21	64.43	23.56	64.43	0.93	18.85	48.96	0.49	12.87	18.42	0.46	10.95	18.42
100	10	63.53	59.06	64.09	59.06	2.54	56.19	39.43	1.08	42.19	10.94	1.14	29.22	10.94
100	15	77.88	61.87	64.12	61.87	2.06	63.73	44.93	0.93	37.41	15.45	0.95	30.26	15.45
100	25	78.34	69.13	64.62	69.13	1.63	43.27	56.49	0.91	31.84	23.35	0.91	28.95	23.35
200	10	440(1)	56.86	1248.75	56.86	26.75	739.05	33.83	9.28	368.87	8.95	9.19	275.96	8.95
200	20	440.19	58.97	436.89	58.97	19.01	267.61	39.59	9.58	118.36	11.04	9.60	82.42	11.04
200	30	349.71	62.75	503.01	62.75	13.78	199.57	46.25	7.80	111.97	15.28	7.84	89.68	15.28

Table 5 reports the computational times and LP gaps for $n \in \{75, 100, 200\}$ if the former preprocessing phase for fixing variables are used in order to reduce the number of variables. The first block of columns corresponds to the formulation without any preprocessing phase and the second one corresponds to the formulation relaxing z_{ir} for $i \in N$, $r = 3, \dots, G_i$. After these two blocks, different options for the preprocessing are studied. In those cases, a first column indicating the preprocessing time is included in each block.

Columns in block “classic rel.” report the results if a preprocessing using (45) and (46) based on the relaxed formulation from Daskin (1995) is used. “Binary” shows the results if Binary algorithm proposed in Calik and Tansel (2013) is used to obtain a lower bound on the p -center for each stratum and the criteria given by (45) and (46) are applied. In columns under heading “Binary*”, the same preprocessing is used but, in this case, z_{ir} variables are relaxed for $i \in N$, $r = 3, \dots, G_i$. The largest differences in CPU time among the variants can be observed in instances with $n = 200$. In this case, the best results regarding CPU time are the ones reported in column “Binary*”. It is worth noting that the preprocessing times represent only a small fraction of the overall solution time in all the instances reported in this

table. Observe also that the LP gaps are considerably reduced if binary algorithm together with (45) and (46) is used.

Table 6: Times of F5-(36)+(43)* using binary algorithm to reduce the number of u -variables and different valid inequalities.

n	p	Binary*	(47)	(48)	(49)	(50)	(51)	(40)
75	5	7.95	12.61	8.02	8.02	7.95	11.95	6.03
75	10	11.04	13.95	11.08	11.11	11.02	16.25	15.90
75	15	10.95	12.96	11.09	11.08	11.01	14.24	9.44
100	10	29.22	37.42	29.21	29.13	29.27	43.85	24.17
100	15	30.26	45.75	30.29	30.28	30.22	46.58	28.28
100	25	28.95	37.38	28.82	28.73	28.76	44.35	36.95
200	10	275.96	539.47	275.58	275.37	276.19	289.21	162.02
200	20	82.42	161.46	82.49	82.69	82.66	92.31	93.18
200	30	89.68	176.37	90.28	90.04	90.00	120.44	164.50
300	15	509.79	1298.54	512.82	513.38	510.10	523.33	271.05
300	30	315.13	591.42	318.61	316.23	316.46	372.64	228.18
300	45	535.69	813.88	538.52	533.47	532.62	442.23	610.77
400	20	1017.28	3305.29	1011.01	1012.90	1014.40	722.30	450.12
400	40	663.16	1863.28	666.45	660.53	663.98	805.02	954.81
400	60	475.14	1246.22	474.36	475.05	474.23	735.84	816.77

Table 6 reports the average times required to solve the same instances with formulation F5-(36)+(43)* using Binary Algorithm, (45) and (46) to reduce the number of variables and adding some of the constraints explained in Subsection 3.5. Regarding the reported results in Table 6, the time performance is significantly improved in some cases if constraints (40) are included as valid inequalities for the formulation. The remaining valid inequalities appearing in this table, except maybe for (47), do not worsen the times in general, but they neither provide a significant improvement.

Finally, Table 7 reports the time results using ORLIB data with the same p values as in the original instances and using random strata. For solving these instances, formulation F5-(36)+(43)* was used with Binary Algorithm and adding (45) and (46) to reduce the number of variables. The results shows that only two instances remain unsolved after two hours using the model with the proposed preprocessing phase (underlined cpu time). In this table, we give separately the time to solve the formulation, under heading t_{solv} , the preprocessing time,

under t_{prep} and the overall time, t_{total} . Additionally, we provide the number of nodes explored in the branch and bound tree. The LP gap is also provided and, in the cases in which the model is not solved in two hours, the LP gap is calculated as $\frac{UB_{\text{best}} - LP}{UB_{\text{best}}} \cdot 100$, where $Best$ is the best objective value obtained in two hours. In these instances (pmed23 and pmed40), the final gap after two hours is shown inside the parentheses in the LP Gap column. Finally, column ‘Obj. Val.’ shows the optimal objective value for each instance except for pmed23 and pmed40 where the best obtained solutions for the unsolved instances in two hours are reported.

In this table we observe that varying p has a strong effect on the CPU times, both, in the preprocessing phase and when solving the final formulation. Moreover, the effect is different in both cases, yielding curious situations, where the preprocessing time can be larger than the actual solution time. We can also observe that the most demanding instances tend to be

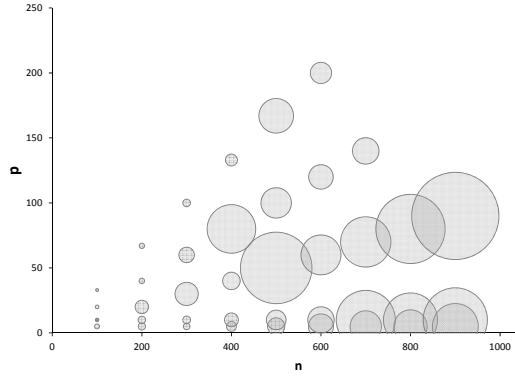


Figure 1: CPU times (circle size) as a function of n and p

those with $p \simeq 10\% \cdot n$. This behavior can be better appreciated in Figure 1.

5.3 SAA for PpCP

In this subsection, the time and gap results of SAA for the PpCP are analyzed. Table 8 shows the results of SAA in comparison with PpCP formulation presented in Martínez-Merino et al. (2017).

The first column corresponds to the running time of the probability chain PpCP formulation described in Martínez-Merino et al. (2017) where we have established a time limit of 24 hours. ‘F1 SAA’ shows the results of SAA if formulation F1 of the SpCP is used. ‘Binary*

SAA” reports again the results of SAA but using formulation F5 with constraints (36) replaced by (43), using Binary Algorithm as a preprocessing phase and relaxing z_{ir} variables for $i \in N$ and $r \in \{3, \dots, G_i\}$. For each block of columns, the gap column reports the gap (in percentage, %) between the best obtained solution in the SAA heuristic and the PpCP objective value. In addition, the time column reports the running time of the procedures.

Regarding the running times of SAA, we observe a significant difference between SAA when using formulation F1 and the remaining SAA columns that use formulation F5. As observed, times in “Binary* SAA” grow much slower than when using F1 so that, even if for the smallest instances they seem to be worse, they become much better for $n > 30$. Considering the gaps we see that in none of the cases, the gaps are bigger than 0.64%. Moreover, both versions of the SAA found the optimal solution for at least half of the instances. As explained in Section 4, we can find theoretical results that guarantee the goodness of the obtained solution when using the SAA.

Table 9 reports the average results of the instances with $(n, p) \in \{(75, 10), (100, 10), (100, 15), (100, 25)\}$. First column reports the necessary time for solving the PpCP using the probability chain formulation, observe that none of the instances were solved in 24 hours. “Gap_{BS}” column reports the gap between the best solution obtained by SAA method and the best solution of PpCP within the time limit. Finally, SAA time is reported. Observe that in all unsolved instances after 24 hours “Gap_{BS}” column reports negative gaps. This is due to the fact that the best solution given by SAA is better than the best solution provided by PpCP formulation after 24 hours.

Table 7: Results for ORLIB data

n		p	solving	prepro	Total	LP Gap	# Nodes	Obj. Val.	n		p	solving	prepro	Total	LP Gap	# Nodes	Obj. Val.
pmed1	100	5	46.00	1.12	47.66	8.75	8041	116.8	pmed21	500	5	120.12	334.01	466.05	4.86	1387	38.9
pmed2	100	10	26.08	1.14	28.13	9.91	4783	95.3	pmed22	500	10	253.44	307.79	597.42	7.78	1713	36.9
pmed3	100	10	11.51	1.09	12.98	8.05	1271	92.2	pmed23	500	50	7200.19	229.07	7447.45	18.36(3.9)	29644	21.6
pmed4	100	20	28.27	0.92	29.48	18.50	9463	72.6	pmed24	500	100	1068.26	320.28	1392.83	22.16	18441	15
pmed5	100	33	18.61	0.72	19.52	37.80	5843	40.6	pmed25	500	167	1591.44	199.79	1793.49	40.16	81497	10.6
pmed6	200	5	75.99	10.02	92.17	6.95	2621	80.5	pmed26	600	5	259.63	677.29	959.90	4.58	1331	36.9
pmed7	200	10	89.48	8.24	99.64	8.59	4737	60.6	pmed27	600	10	427.28	593.22	1056.32	8.91	2315	31.1
pmed8	200	20	279.66	8.63	290.36	13.40	10471	52.4	pmed28	600	60	1784.88	572.43	2371.33	13.53	6377	17.3
pmed9	200	40	48.49	8.75	57.95	17.09	4737	34.5	pmed29	600	120	487.86	425.46	918.56	22.71	10189	12
pmed10	200	67	47.67	6.26	54.21	33.24	10931	18.3	pmed30	600	200	188.01	505.15	696.21	32.72	5877	9
pmed11	300	5	30.98	44.26	78.77	5.05	353	54.6	pmed31	700	5	132.35	1318.46	1470.65	2.73	393	29.2
pmed12	300	10	57.64	41.64	105.20	6.77	1215	50.9	pmed32	700	10	4250.54	788.62	5104.98	7.09	19239	28
pmed13	300	30	786.58	33.21	829.07	17.75	13347	35.1	pmed33	700	70	2666.15	1024.91	3729.87	13.66	7367	14.9
pmed14	300	60	338.51	38.84	380.14	19.62	9919	23.4	pmed34	700	140	305.98	752.26	1067.68	16.55	3559	10.3
pmed15	300	100	60.15	40.86	101.79	27.72	6159	16.1	pmed35	800	5	180.27	1429.91	1647.11	4.18	935	28.5
pmed16	400	5	47.96	114.38	165.70	3.88	599	45.9	pmed36	800	10	2376.87	1837.73	4283.38	6.03	4411	26.4
pmed17	400	10	184.14	115.38	309.41	6.19	2399	38.5	pmed37	800	80	5553.53	1353.61	6987.60	14.52	9285	14.7
pmed18	400	40	384.20	86.96	480.31	9.80	4275	26.7	pmed38	900	5	207.01	2844.10	3109.63	5.43	289	27.5
pmed19	400	80	3354.26	119.21	3476.32	21.85	64273	17.4	pmed39	900	10	2993.05	2981.84	6021.92	7.82	6587	22.9
pmed20	400	133	137.76	93.43	232.63	39.05	10437	12.7	pmed40	900	90	7205.28	3780.23	11033.40	13.07(3.9)	10222	12.7

Table 8: SAA results

n	p	PpCP	F1 SAA		Binary* SAA	
		Time	Gap	Time	Gap	Time
6	2	0.01	0.00	0.10	0.00	0.45
10	3	0.03	0.00	0.54	0.00	0.87
10	5	0.03	0.00	0.49	0.03	1.49
13	3	0.07	0.00	0.91	0.00	1.53
13	5	0.07	0.00	1.10	0.24	2.21
13	8	0.05	0.00	1.32	0.00	2.20
15	3	0.10	0.00	1.77	0.00	1.88
15	7	0.13	0.00	1.73	0.10	3.56
15	10	0.07	0.00	1.84	0.56	2.65
20	3	0.32	0.00	3.40	0.00	4.18
20	7	0.63	0.64	3.67	0.00	8.49
20	10	0.49	0.10	4.28	0.14	5.62
25	3	0.84	0.00	6.73	0.00	7.42
25	7	3.48	0.05	8.96	0.23	9.50
25	10	5.13	0.02	9.48	0.01	14.34
30	3	2.01	0.00	13.90	0.00	11.27
30	7	13.61	0.14	12.78	0.15	9.40
30	10	22.99	0.00	16.24	0.00	16.54
40	3	8.28	0.00	40.90	0.00	19.94
40	7	148.22	0.01	98.39	0.20	19.45
40	10	295.52	0.01	96.68	0.01	19.52
50	5	243.17	0.03	162.76	0.00	44.68
50	10	4083.75	0.01	462.26	0.12	67.74
50	15	21782.53	0.21	794.07	0.01	71.63
75	5	4108.22	0.03	1386.77	0.03	150.28

Table 9: SAA results for larger instances.

n	p	PpCP Time	Gap _{BS}	SAA Time
75	10	> 86400	-2.55	200.57
75	15	> 86400	-5.67	258.32
100	10	> 86400	-10.25	491.73
100	15	> 86400	-15.31	449.02
100	25	> 86400	-20.40	850.55

6 Conclusions

This paper presents an extension of the p -center problem called the Stratified p -Center Problem ($SpCP$). This extension could be applied in cases where the population is divided into different strata and the evaluation of the service must be separately measured for each stratum. In the model, it is assumed that more than one stratum can be present at each demand point.

Different formulations were introduced together with a detailed study of variants, variable reduction processes and valid inequalities. Regarding the computational results, the best performance was obtained using a formulation based on covering variables.

The $SpCP$ allows to implement a heuristic approach based on the Sample Average Approximation (SAA) method to obtain good feasible solutions for the probabilistic p -center problem. This heuristic approach provides good upper bounds in acceptable times.

Acknowledgements

A.M. Rodríguez-Chía and Luisa I. Martínez Merino acknowledge that research reported here was supported by the European Regional Development's funds (FEDER) and Agencia Estatal de Investigación (AEI) under project MTM2016-74983-C2-2-R. Luisa I. Martínez Merino was also supported by Universidad de Cádiz PhD grant UCA/REC02VIT/2014 and Programa de Fomento e Impulso de la actividad Investigadora UCA (2018). The research of Maria Albareda has been partially funded by Spanish Ministry of Economy and Competitiveness and EDRF funds through project MTM2015-63779-R.

References

- M. Albareda-Sambola, J. Díaz, and E. Fernández. Lagrangean duals and exact solution to the capacitated p-center problem. *European Journal of Operational Research*, 201(1): 71–81, 2010.
- I. Averbakh and O. Berman. Minimax regret p-center location on a network with demand uncertainty. *Location Science*, 5(4):247–254, 1997.
- M. Balinski. Integer programming: methods uses, computation. *Management Science*, 12: 253–313, 1965.
- H. Calik and B. C. Tansel. Double bound method for solving the p-center location problem. *Computers and Operations Research*, 40(12):2991–2999, 2013.
- H. Calik, M. Labbé, and H. Yaman. p-center problems. In *Location Science*, pages 79–92. Springer, 2015. ISBN 978-3-319-13110-8.
- B. Callaghan, S. Salhi, and G. Nagy. Speeding up the optimal method of Drezner for the p-centre problem in the plane. *European Journal of Operational Research*, 257(3):722–734, 2017.
- D. Chen and R. Chen. Optimal algorithms for the α -neighbor p-center problem. *European Journal of Operational Research*, 225(1):36–43, 2013.
- C. Contardo, M. Iori, and R. Kramer. A scalable exact algorithm for the vertex p-center problem. *Computers & Operations Research*, 103:211 – 220, 2019.
- M. Daskin. *Network and Discrete Location: Models, Algorithms, and Applications*. Wiley, New York, 1995.
- G. Dobson and U. Karmarkar. Competitive location on a network. *Operational Research*, 35: 565–574, 1987.
- Z. Drezner. Conditional p-center problems. *Transportation Science*, 23(1):51–53, 1989.
- S. Elloumi, M. Labbé, and Y. Pochet. A new formulation and resolution method for the p-center problem. *INFORMS Journal on Computing*, 16(1):84–94, 2004.

- A. Elshaikh, S. Salhi, J. Brimberg, N. Mladenović, B. Callaghan, and G. Nagy. An Adaptive perturbation-based heuristic: An application to the continuous p -centre problem. *Computers & Operations Research*, 75:1–11, 2016.
- I. Espejo, A. Marín, and A. M. Rodríguez-Chía. Closest assignment constraints in discrete location problems. *European Journal Operational Research*, 219:49–58, 2012.
- I. Espejo, A. Marín, and A. M. Rodríguez-Chía. Capacitated p -center problem with failure foresight. *European Journal of Operational Research*, 247(1):229–244, 2015.
- S. García, M. Labbé, and A. Marín. Solving large p -median problems with a radius formulation. *INFORMS Journal on Computing*, 23(4):546–556, 2011.
- R. Garfinkel, A. Neebe, and M. Rao. The m -center problem: Minimax facility location. *Management Science*, 23(10):1133–1142, 1977.
- T. Homem-de-Mello and G. Bayraksan. Monte carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56–85, 2014.
- C. Irawan, S. Salhi, and Z. Drezner. Hybrid meta-heuristics with VNS and exact methods: application to large unconditional and conditional vertex p -centre problems. *Journal of heuristics*, 22(4):507–537, 2016.
- M. Jeger and O. Kariv. Algorithms for finding p -centers on a weighted tree (for relatively small p). *Networks*, 15(3):381–389, 1985.
- O. Kariv and S. Hakimi. An algorithmic approach to network location problems I: The p -centers. *SIAM Journal on Applied Mathematics*, 37(3):513–538, 1979.
- A. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.
- J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142(1):215–241, 2006.

- C.-C. Lu and J.-B. Sheu. Robust vertex p-center model for locating urgent relief distribution centers. *Computers & Operations Research*, 40(8):2128–2137, 2013.
- A. Marín, S. Nickel, J. Puerto, and S. Velten. A flexible model and efficient solution strategies for discrete location problems. *Discrete Applied Mathematics*, 157(5):1128–1145, 2009.
- L. I. Martínez-Merino, M. Albareda-Sambola, and A. M. Rodríguez-Chía. The probabilistic p-center problem: Planning service for potential customers. *European Journal of Operational Research*, 262(2):509–520, 2017.
- F. A. Özsoy and M. Ç. Pınar. An exact algorithm for the capacitated vertex p-center problem. *Computers & Operations Research*, 33(5):1420–1436, 2006.
- D. R. Quevedo-Orozco and R. Z. Ríos-Mercado. Improving the quality of heuristic solutions for the capacitated vertex p-center problem through iterated greedy local search with variable neighborhood descent. *Computers & Operations Research*, 62:133–144, 2015.
- C. Revelle and K. Hogan. The maximum reliability location problem and α -reliable-p-center problem: Derivatives of the probabilistic location set covering problem. *Annals of Operations Research*, 18(1):155–173, 1989.
- S. M. Robinson. Analysis of sample-path optimization. *Mathematics of Operations Research*, 21(3):513–528, 1996.
- R. Y. Rubinstein and A. Shapiro. Optimization of static simulation models by the score function method. *Mathematics and Computers in Simulation*, 32(4):373–392, 1990.
- D. Schilling, D. J. Elzinga, J. Cohon, R. Church, and C. ReVelle. The team/fleet models for simultaneous facility and equipment siting. *Transportation Science*, 13(2):163–175, 1979.
- A. Shapiro. Sample average approximation. In *Encyclopedia of Operations Research and Management Science*, pages 1350–1355. Springer, 2013.