

A New Generalization to the Capacitated Centred Clustering Problem

Marcos Negreiros

Universidade Estadual do Ceará
Av. Silas Munguba, 1700 - Itaperi, Fortaleza/CE, CEP: 60.714-903
marcos.negreiros@uece.br

Nelson Maculan

Universidade Federal do Rio de Janeiro,
Programa de Eng de Sistemas e Computação (PESC-COPPE),
Ilha do Fundão, Sala H-322B, CEP: 21.941-972, Rio de Janeiro, RJ, Brazil
maculan@cos.ufrj.br

Renan Vicente Pinto

Universidade Federal Rural do Rio de Janeiro
renanvp@ufrj.br

Philippe Michelon

Université de Avignon
philippe.michelon@univ-avignon.fr

RESUMO

Este artigo propõe uma nova extensão para o problema de agrupamento capacitado em centro geométrico, denominada problema periódico de agrupamento capacitado, onde são dados itens com atributos de localização, demanda, periodicidade de cobertura, grupos limitados com capacidade heterogênea e custo fixo associado a cada grupo. Trata-se de uma versão rica de cobertura capacitada. Deseja-se definir os agrupamentos que minimizam a dissimilaridade entre itens do mesmo grupo de modo que não ultrapassem a sua demanda máxima de formação definida e a periodicidade entre coberturas dos itens. Apresentamos uma formulação geral e sua versão quadrática para o problema, com novo conjunto de restrições periódicas, que facilitam a resolução do problema por resolvedores como Gurobi, CPLEX e XPress. Apresentamos resultados sobre instâncias testes e reais da coleta de lixo hospitalar com até 78 itens, 6 dias de período e 8 grupos heterogêneos e os comparamos usando os solvers indicados.

PALAVRAS CHAVE. Clustering, Periodic Clustering, Combinatorial Optimization.

OC, PM

ABSTRACT

This article proposes a new generalization to the capacitated clustering problem in geometric center, called periodic capacitated clustering problem, where items with attributes of location, demand, coverage periodicity, limited to a set of heterogeneous capacitated clusters, associated fixed cost of a given cluster capacity and others that characterize them are given. It is desired to define the groupings that minimize the dissimilarity between items in the same group so that they do not exceed the maximum demand of the group and the periodicity between coverage. We present

a general formulation and its quadratic version for the problem, with second order cone constraints, which facilitate solving the problem by solvers such as Gurobi, CPLEX and XPress. We show results for test and real instances from health waste collection with up to 78 items, 6 days in a time horizon and 8 heterogeneous clusters and we compare the results obtained between the solvers.

KEYWORDS. Clustering. Periodic Clustering. Combinatorial Optimization.

OC, PM

1. Introduction

In medical or health waste collection system performed in major Brazilian cities, the garbage is collected in health units that are: major hospitals, UPAs, health centers, and clinics that produces hazardous and/or contagious organic and solid waste materials, Negreiros [1994], Chartier et al. [2014], Taslimi et al. [2020], UNDPGEF [2023]. Mainly the health units have separated hazardous and non-hazardous garbage that are both disposed in a autoclave/incinerator final destination place. The production of waste per health unity is non-deterministic daily, and it is common that the items are devised in ABC standard classification of production and degree of dangerousness to define the basic periodic time of the item to be serviced, Chartier et al. [2014]. This type of waste may not be stored for long ($< 72h$), and it is common that the maximum horizon is among 1-6 days of visit. Careful knowledge about the waste demand of each unity has to be taken, Negreiros [1994], Taslimi et al. [2020]. In the other side fleet is in general very specific with different capacity and resources in safety and stored tank to avoid melting garbage. The reverse logistic transportation system can be designed by considering all these factors, to maximize coverage at minimum cost of resources and assignment to the fleet.

A solution approach using the CCCP and after applying VRP tools was proposed by Araujo et al. [2013]. The authors consider a type of cluster-first-route-second to the homogeneous periodic VRP using the cluster search metaheuristic with close results to the best upper-bound already known. The authors applied with success their methodology for a instance produced by Bianchi-Aguiar et al. [2012] with 51 items and 6 periods from waste collection of the city of Ponte de Lima/PT. As far as we know, no other reference in the literature consider homogeneous or heterogeneous clustering incorporated as first phase to solve PVRP in the health waste collection case.

In other hand, Negreiros et al. [2022] proposed a methodology for a mathematical programming based decision support system (DSS) - **SisRot®Full** to deal with this reverse logistic system.

A single period (i.e. one day) appropriate model to solve this type of problem was proposed by Negreiros e Palhano [2006] by using the capacitated centred clustering problem (CCCP). The general version ($gCCCP$) is more appropriate when the health units need to be known, and the single version ($pCCCP$) is more appropriate when the loadpackers' fleet composition is known in advance.

In other direction, ideas for clustering problems can be directly obtained from the periodic VRP, because it is one of the most studied problems in combinatorics with the periodic component. The problem was proposed by Beltrami e Bodin [1974] for waste collection and is being studied by a number of researchers, Gaudioso e Palleta [1992], Cordeau et al. [1997], Francis e K. [2006]. One of the most referred periodic VRP formulations consider the use of patterns of coverage between assignments accordingly customers demands per period. Solutions based on this formulation were proposed by Tan e Beasley [1984] profiting the results obtained with the generalized assignment VRP method developed by Fisher e Jaikumar [1981], Fisher et al. [1982].

Exact formulations for the PVRP were tested formally by Foster e Ryan [1976], and the best IP models were produced by Baldacci et al. [2011] with gaps less than 1% from the optimality for the traditional 32 PVRP ("r") instances. Most recently Huerta-Munoz et al. [2022] proposed three models to solve the Flexible Periodic VRP, incorporating periodic constraints with no patterns. A literature review on this problem was made by Campbell e Wilson [2014] celebrating 20 years of the studies in the PVRP.

Approaches based on periodic vehicle routing with inventory component were investigated by a number of researchers for medical waste collection, see Taslimi et al. [2020].

In the literature there are other important direction called *Districing*. It is the problem of grouping small geographic areas, into larger geographic clusters(districts), such that the latter are balanced, contiguous, and compact. Balance describes the desire for districts of equitable size. A district is said to be geographically compact if it is somewhat round-shaped and undistorted. Districing problems are motivated by quite different applications ranging from political districting over the design of districts for schools, social facilities, waste collection, or winter services, to sales and service territory design. Despite the considerable number of publications on districing problems, there is no consensus on which criteria are eligible and important and, moreover, on how to measure them appropriately, for further reading about districing the authors refer to Kalcsics e Rios-Marcado [2019], Rios-Marcado [2020]. Districing applied to the design of general practitioners (GP) territories sites in Turkey were developed by Yanik et al. [2019]. Recent approach for emergency medical (EMS) system was proposed by ?. Promote districing with capacity and periodic constraints is the main goal of this paper.

In the original single period *g*CCCP model proposed by Negreiros e Palhano [2006], the capacity of the *clusters* are homogeneous, and the objective function considers the minimization of the total variance between items and their *cluster* center. Also, it is considered non-linear constraints to warranty the geometric centers of the clusters. As can be seen below, we consider a new formulation where the *clusters* of distinct capacities (i.e. related to each salesman daily sell capacity), we remove the center constraints and maintain the minimization of the fixed cost of opening a new cluster summed up to the total variance (dissimilarity) between items and assigned *cluster* centers. For this case the problem can be represented by using the following sets, parameters and variables:

Sets r is the dimension of the space ($r = 2$, in this case); I is the set of individuals; J is the set of *clusters* centers; and $|J|$ is the cardinality of the set J , or a fixed number of *clusters* ($|J| = p$).

Parameters p is the number of *clusters*; a_i is an r -dimensional vector (related to the positional and characteristics' attributes of the individuals); α_j is a positive factor associated with variable cost of including cluster of j capacity; q_i is the demand of an individual i and Q_j is the maximum capacity of a *cluster*; F_j is the fixed cost associated on opening a new cluster j .

Variables $\bar{x}_j$ is an r -dimensional vector represented by the center coordinates of *cluster* j ; $z_j = 1$ if cluster j is open and 0 otherwise; and $y_{ij} = 1$ if an individual i is assigned to the *cluster* j or $y_{ij} = 0$, otherwise.

A formulation for the general heterogeneous capacitated centred clustering problem (*g*HCCCP) can be done as following, as proposed by Negreiros et al. [2020] with minor extension,

$$(\text{gHCCCP}) \text{ Minimize } \sum_{j \in J} F_j z_j + \sum_{j \in J} \alpha_j \sum_{i \in I} \|a_i - \bar{x}_j\|^2 y_{ij} \quad (1)$$

$$\text{such that : } \sum_{j \in J} y_{ij} = 1, \quad i \in I \quad (2)$$

$$\sum_{i \in I} q_i y_{ij} \leq Q_j z_j, \quad j \in J \quad (3)$$

$$\bar{x}_j \in \mathbb{R}^r, j \in J \quad (4)$$

$$y_{ij} \in \{0, 1\}, \quad i \in I, j \in J \quad (5)$$

$$z_j \in \{0, 1\}, \quad j \in J \quad (6)$$

The objective function (1) minimizes the Euclidean distance between *clusters* centers and their assigned items. The constraint (2) assigns one individual to just one *cluster*. The constraint (3) limits the assigned individuals to the maximum capacity of the *cluster* j . The constraints (4)-(6) refer to the decision variables of the problem.

$g\text{HCCCP}$ is non-linear, non-convex and mixed-binary. It is NP-Hard as the previous problems, because its unconstrained version is also NP-Hard, Negreiros e Palhano [2006]. It seems that the $g\text{HCCCP}$ problem is further more difficult to be solved than $g\text{CCCP}$. The formulation can return quality solutions of any kind, being the set of items disperse or not. In figures 1 and 2 (source: ?) are presented an example of an instance where items have unitary demand and the optimal solution of the $g\text{HCCCP}$ respectively.

This model is $O(I \cdot J)$ in the number of binary variables and $O(I)$ in the number of constraints. It is necessary a non-linear- $\{0, 1\}$ solver to run this model. Solvers like FilMINT, Knitro and BARON from NEOS Server, Czyzyk et al. [1998], Dolan [2001] and Gropp e More [1997], and LINGO from LINDO Systems, can only solve this model close to optimality for some very small instances, see Negreiros et al. [2020].

In the following section 2, to design periodic district areas composed by their clustered items, we propose new models for the periodic fixed cost capacitated centred clustering problem in the sub-section 2.1, that consider individuals with same demands per unit of the time horizon, and in the sub-section 2.2 we propose a quadratic generalized formulation considering only individuals with periodic circles and distinct demands per time horizon. We show the problem by solving an

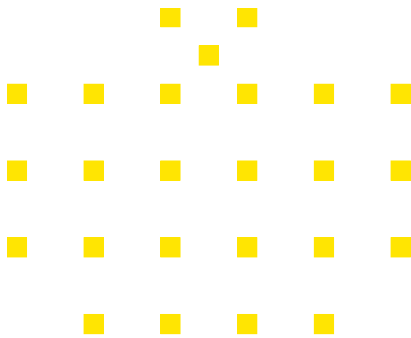


Figure 1: Heterogeneous Capacitated Clustering instance.

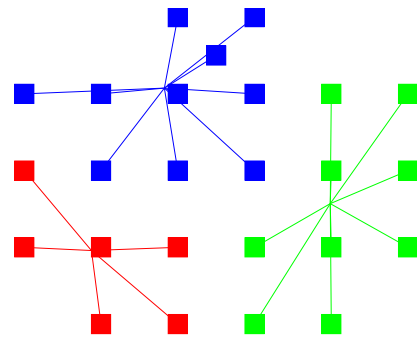


Figure 2: $g\text{HCCCP}$ associated optimal solution.

example to better qualify its response. In section 3 we test our models for the generalized periodic capacitated centred clustering problem (*gPHCCCP*) instances from synthetic and real-life using state-of-the-art MILP-quadratic solvers.

2. Periodic Heterogeneous Capacitated Centred Clustering Problem

It can be proposed two different models for clustering individuals. In the first approach, it is considered that the demand of the individuals is given for each day of the time horizon. One can consider that the pattern of periodic coverage of each individual is previously known and may be fixed for some/all of them, as in the formulation of the PVRP. In the objective the clusters fixed cost are to be minimized as the total assignment distance between their individuals and the centers. This problem we call - Periodic Fixed Coverage Heterogeneous Capacitated Centred Clustering Problem (PFC-HCCCP)

In the second approach we consider that a set of individuals can be periodically serviced accordingly the individual self periodicity pattern defined by the minimum and maximum times in the period it has to be covered. The demand of the individuals can change along the period, and the clusters' fixed added to variable costs related to the total variance to assign items to centers along the period are to be minimized. This problem we call - Generalized Periodic Heterogeneous Capacitated Centred Clustering Problem (*gPHCCCP*).

PFC-HCCCP and *gPHCCCP* are problems that can be applied in wholesales distribution, waste collection (domiciliary, health, commercial, community, street-sweeping), epidemiologic prevention and combat, utility load maneuvers and maintenance, road maintenance, aggro-food supply chain network, vehicle routing, bulk milk collection and many other, Negreiros e Palhano [2006], Negreiros et al. [2011], Bard e Jarrah [2013], Negreiros et al. [2022].

2.1. PFC-HCCCP

Let us suppose that the periodic fixed coverage capacitated centred clustering problem (PFC-HCCCP) can be represented using the following set of parameters and variables:

l : dimension of the Euclidean space ($l = 2$, in this case).

L : set of dimensions of the Euclidean space ($|L| = l$, in this case).

I : set of individuals.

J : set of clusters.

H : set of time horizon, $H = \{h_1, h_2, \dots, h_T\}$.

P : set of patterns of periodic coverage in time horizon H .

a_i : vector of l dimension with the coordinates of the individual i .

T : period of time horizon.

d_h^p : if time h is in pattern p .

FC_j : is the fixed cost associated on opening a new cluster j .

α_j : is a positive factor associated with the variable cost to use cluster of j capacity.

Q_j : maximum capacity of a cluster j for any time horizon.

q_i^h : demand of an individual i in time horizon h .

$$z_j = \begin{cases} 1, & \text{if cluster } j \text{ is assigned} \\ 0, & \text{otherwise} \end{cases}$$

$$u_{ih}^p = \begin{cases} 1, & \text{if individual } i \text{ is assigned to time horizon } h \text{ of pattern } p \\ 0, & \text{otherwise} \end{cases}$$

$\bar{x}_j^h$: vector of dimension l representing the coordinates of the *cluster* center j in time horizon h .

$$y_{ij}^h = \begin{cases} 1, & \text{if individual } i \text{ is assigned to cluster } j \text{ in time horizon } h \\ 0, & \text{otherwise} \end{cases}$$

We consider the periodic fixed cost heterogeneous capacitated centred clustering problem (PFC-HCCCP) as follows:

$$\text{(PFC-HCCCP) Minimize } \sum_{j \in J} FC_j z_j + \sum_{h \in H} \sum_{j \in J} \alpha_j \sum_{i \in I} \|a_i - \bar{x}_j^h\| y_{ij}^h, \quad (7)$$

$$\text{such that: } \sum_{j \in J} y_{ij}^h = 1, i \in I, h \in H, \quad (8)$$

$$\sum_{p \in P} u_{ih}^p = 1, i \in I, h \in H \quad (9)$$

$$\sum_{j \in J} y_{ij}^h = u_{ih}^p d_h^p, i \in I, h \in H, p \in P, \quad (10)$$

$$\sum_{i \in I} q_i^h d_h^p y_{ij}^h \leq Q_j z_j, j \in J, h \in H, \quad (11)$$

$$\bar{x}_j^h \in \mathbb{R}^l, y_{ij}^h \in \{0, 1\}, i \in I, j \in J, h \in H, \quad (12)$$

$$z_j \in \{0, 1\}, u_{ih}^p \in \{0, 1\}, i \in I, j \in J, h \in H. \quad (13)$$

The objective function (7) minimizes the fixed cost to open a new cluster j and the total variance among individuals assigned to the same cluster center j into period h . Constraint (8) assigns each individual to exactly one cluster in time horizon h . Constraints (9-10) refer to the assignment of the items' best periodic pattern assignment to clusters. For each cluster center j in time horizon h , constraint (11) limits the sum of the demands of its assigned items. Constraint (12-13) refers to the domain of the decision variables.

The PFC-HCCCP is non-linear with linear constraints (8-11). Its difficulty (NP-Hardness) can be found in location of the cluster centers, knapsack constraint in (11), the periodic component in (9-10), and the partitioning process encapsulated in the problem.

The periodical constraints can be changed by moving the set of periods to related constraints that represent a set of equations determining the feasible cycles directly selected by the model, as is exposed further in equations (31-32) in the model (gPHCCCP).

A continuous relaxation for PFC-HCCCP, whose objective function is neither convex nor differentiable, can be modified to be differentiable and convex by applying second-order cone constraints, Boyd e Vandenberghe [2004], Lobo et al. [2019] and Fawzi [2019], as proposed for clustering problems by Negreiros et al. [2020] and Linhares et al. [2022].

Therefore, the problem PFC-HCCCP can be rewritten as follows.

$$(PFC-HCCCP') \text{ Minimize } \sum_{j \in J} FC_j z_j + \sum_{i \in I} \sum_{j \in J} \sum_{h \in H} w_{ij}^h \quad (14)$$

Such that,

$$(w_{ij}^h)^2 \geq \alpha_j \sum_{r=1}^l (m_{ijr}^h)^2, \quad w_{ij}^h \geq 0, \quad i \in I, j \in J, h \in H, \quad (15)$$

$$-M(1 - y_{ij}^h) + (a_{ir} - \bar{x}_{jr}^h) \leq m_{ijr}^h, \quad r \in L, i \in I, j \in J, h \in H, \quad (16)$$

$$m_{ijr}^h \leq (a_{ir} - \bar{x}_{jr}^h) + M(1 - y_{ij}^h), \quad r \in L, i \in I, j \in J, h \in H, \quad (17)$$

$$-M y_{ij}^h \leq m_{ijr}^h \leq M y_{ij}^h, \quad r \in L, i \in I, j \in J, h \in H, \quad (18)$$

$$\sum_{j \in J} y_{ij}^h \geq 1, \quad i \in I, h \in H, \quad (19)$$

$$\sum_{i \in I} q_i^h d_h^p y_{ij}^h \leq Q_j z_j, \quad j \in J, h \in H, \quad (20)$$

$$\sum_{p \in P} u_i^p = 1, \quad i \in I \quad (21)$$

$$u_i^p d_h^p = \sum_{j \in J} y_{ij}^h, \quad i \in I, h \in H, p \in P \quad (22)$$

$$\bar{x}_{jr}^h = (\bar{x}_{j1}^h \bar{x}_{j2}^h \dots \bar{x}_{jl}^h)^\top \in \mathbb{R}^l, \quad j \in J, h \in H \quad (23)$$

$$u_i^p \in \{0, 1\}, \quad i \in I, p \in P \quad (24)$$

$$y_{ij}^h \in \{0, 1\}, \quad i \in I, j \in J, h \in H \quad (25)$$

$$m_{ijr}^h \in \mathbb{R}, \quad i \in I, j \in J, h \in H, r \in L \quad (26)$$

$$a_{ir} = (a_{i1} a_{i2} \dots a_{il})^\top, \quad (27)$$

$$z_j^h \in \{0, 1\}, \quad j \in J, h \in H \quad (28)$$

Constraints (15) are the second-order cones, which form a convex set. The PFC-HCCCP' is a quadratic problem with quadratic-binary-linear constraints. If the vector of periodic assignments, u_{ip}^h , here used as in the PVRP formulations, does not reflect any combinatorial explosion of alternatives, this new formulation is $O(J L H)$ in the number of variables and $O(J L H)$ in the number of constraints, turning the problem to be considered by most quadratic binary-linear solvers like Gurobi, CPLEX and XPress. But it may explode if the period grows and the alternatives to cover it are high.

2.2. gPHCCCP

Let us suppose that the generalized periodic heterogeneous capacitated centred clustering problem (gPHCCCP) can be represented using the following set of parameters and variables:

l : dimension of the Euclidean space ($l = 2$, in this case).

L : set of dimensions of the Euclidean space ($|L| = 2$, in this case).

I : set of individuals.

J : set of clusters.

H : set of time horizon, $H = \{h_1, h_2, \dots, h_T\}$.

a_i : vector of l dimension with the coordinates of the individual i .

FC_j : is the fixed cost associated on opening a new cluster j .

α_j : is a positive factor applied to the variable cost of using cluster of j capacity.

T : period of time horizon.

LI_i : minimum time between two visits of an item i during time horizon.

LF_i : maximum time between two visits of an item i during time horizon.

Q_j : maximum capacity of a cluster j .

q_i^h : demand of an individual i for time horizon h .

$\bar{x}_j^h$: vector of dimension l representing the coordinates of the *cluster* center j in time horizon h .

$$y_{ij}^h = \begin{cases} 1, & \text{if individual } i \text{ is assigned to cluster } j \text{ in time horizon } h \\ 0, & \text{otherwise} \end{cases}$$

$$zv_j = \begin{cases} 1, & \text{if cluster } j \text{ is used} \\ 0, & \text{otherwise} \end{cases}$$

$$zz_i^h = \begin{cases} 1, & \text{if individual } i \text{ is assigned to time horizon } h \\ 0, & \text{otherwise} \end{cases}$$

We consider the generalized periodic heterogeneous capacitated centred clustering problem $gPHCCCP$ as follows:

$$(gPHCCCP) \text{ Minimize } \sum_{j \in J} FC_j z v_j + \sum_{h \in H} \sum_{j \in J} \alpha_j \sum_{i \in I} \|a_i - \bar{x}_j^h\|^2 y_{ij}^h, \quad (29)$$

$$\text{such that:} \quad (30)$$

$$\sum_{j \in J} \sum_{h \in H} y_{ij}^h \geq 1, i \in I, \quad (31)$$

$$\sum_{h \in H} y_{ij}^h \geq z v_j, i \in I, j \in J, \quad (32)$$

$$\sum_{i \in I} y_{ij}^h \geq z v_j, j \in J, h \in H, \quad (33)$$

$$\sum_{i \in I} q_i y_{ij}^h \leq Q_j z v_j, j \in J, h \in H, \quad (34)$$

$$\sum_{h=1}^{LF_i} z z_i^h \geq 1, i \in I \quad (35)$$

$$\sum_{j \in J} y_{ij}^h = z z_i^h, i \in I, h \in H, \quad (36)$$

$$\sum_{s=LI_i}^{LF_i} z z_i^{h+s} \geq z z_i^h, i \in I, h = 1 \dots |H| - LF_i, \quad (37)$$

$$\bar{x}_j^h \in \mathbb{R}^l, y_{ij}^h \in \{0, 1\}, z z_i^h \in \{0, 1\}, i \in I, j \in J, h \in H, \quad (38)$$

$$z v_j \in \{0, 1\}, j \in J, h \in H \quad (39)$$

The objective function (29) minimizes the total variance among individuals assigned to the same cluster in horizon-day h and the total fixed cost to use selected clusters with known maximum capacity (It is also possible to use Fermat-Weber distance - Euclidean distance, as variable cost). Constraint (31) defines the assignment of an individual to at least one cluster in any time horizon. Constraint (32) assigns an individual to a cluster if the selected cluster type j is used in any time horizon h . Constraint (33) indicates that if a cluster type j is used at least one individual is assigned in any time horizon h . For each cluster center j in time horizon h , constraint (34) limits the sum of the demands of its assigned items. Constraint (35) guarantee that an item will be covered until its initial time limit LI . Constraint (36) selects the cyclic period of individual i ($z z_i^h$) in time horizon h . Constraint (37) warranties the cyclic behavior of individual i along time horizon accordingly its limits between two consecutive visits. Constraints (38-39) refers to the domain of the decision variables.

Constraints (34-36) were proposed by Negreiros et al. [2011], that guarantees the cyclic coverage of the zones, avoiding the combinatorial explosion of alternatives as the ones used from the PVRP formulations. These constraints reduces the number of variables needed to solve the periodic problem.

The $gPHCCCP$ problem is non-linear and non-differentiable with linear constraints as it is. But, it can be easily turned to a sequential quadratic problem, by changing the objective function and incorporating convex and quadratic constraints (15-18) of PFC-HCCCP' above to the problem. This results in the following formulation:

$$(gPHCCCP') \text{ Minimize } \sum_{j \in J} FC_j z v_j + \sum_{h \in H} \sum_{j \in J} \sum_{i \in I} w_{ij}^h, \quad (40)$$

Such that,

$$(w_{ij}^h) \geq \alpha_j \sum_{r=1}^l (m_{ijr}^h)^2, \quad w_{ij}^h \geq 0, \quad i \in I, j \in J, h \in H, \quad (41)$$

$$-M(1 - y_{ij}^h) + (a_{ir} - \bar{x}_{jr}^h) \leq m_{ijr}^h, r \in L, i \in I, j \in J, h \in H, \quad (42)$$

$$m_{ijr}^h \leq (a_{ir} - \bar{x}_{jr}^h) + M(1 - y_{ij}^h), r \in L, i \in I, j \in J, h \in H, \quad (43)$$

$$-M y_{ij}^h \leq m_{ijr}^h \leq M y_{ij}^h, r \in L, i \in I, j \in J, h \in H, \quad (44)$$

$$\sum_{j \in J} \sum_{h \in H} y_{ij}^h \geq 1, i \in I, \quad (45)$$

$$\sum_{h \in H} y_{ij}^h \geq z v_j, i \in I, j \in J, \quad (46)$$

$$\sum_{i \in I} y_{ij}^h \geq z v_j, j \in J, h \in H, \quad (47)$$

$$\sum_{i \in I} q_i y_{ij}^h \leq Q_j z v_j, j \in J, h \in H, \quad (48)$$

$$\sum_{h=1}^{LF_i} z z_i^h \geq 1, i \in I \quad (49)$$

$$\sum_{j \in J} y_{ij}^h = z z_i^h, i \in I, h \in H, \quad (50)$$

$$\sum_{s=LF_i}^{LF_i} z z_i^{h+s} \geq z z_i^h, i \in I, h = 1 \dots |H| - LF_i, \quad (51)$$

$$\bar{x}_j^h \in \mathbb{R}^l, y_{ij}^h \in \{0, 1\}, z z_i^h \in \{0, 1\}, i \in I, j \in J, h \in H \quad (52)$$

$$z v_j \in \{0, 1\}, j \in J \quad (53)$$

The formulation above may be extended to include items with fixed period of coverage by just including the constraints (10) in model gPHCCCP', considering only those items with defined fixed period. To consider the clusters as homogeneous (same maximum capacity) it is enough to change equation 34 to hold the fixed capacity, and also altering the demand to be fixed along periods.

As an example, let us consider the instance as shown in Figure 3 where there are 10 items in the Euclidean space, and in Table 2 we have the data associated with each item. The items are shaped in different forms (square and triangle) to differentiate the ones that have a diary frequency from those that do not. We have the possibility to open 3 groups with capacity and associated fixed cost equal to $\{(Q_j, CF_j) : (10, 10k), (12, 20k), (15, 40k)\}$.

The instance indicates the different demands per item during the days, also it maintains the capacity of the clusters along the period as the same. This consideration makes the clusters of the solution fluctuate along the area, Figure 4, once they have to be present in each day. For each day h of the period the total load Qt of the formed clusters are $\{(h : Qt_1, Qt_2, Qt_3) : (1 : 9, 2, 6), (2 : 8, 9, 14), (3 : 1, 4, 15)\}$. In this solution the interesting part is the comparison between

two different days (1st and 3rd), it can be seen that just one item with non diary frequency is used and the clusters, although with the same capacity they changed considerably. This is the optimal solution and it returns as fixed cost equal 70000.00 and variable cost equal 318333.33 (Objective Function = 388333.33). This instance is related to 2Cycle03 whose result is reported in Table 6 by CPLEX.

Item	a_{i1}	a_{i2}	q_i^1	q_i^2	q_i^3	LI_i	LF_i
1	100	500	2	5	1	1	1
2	500	400	2	5	4	1	1
3	300	300	3	7	1	1	1
4	500	100	1	2	6	1	1
5	200	100	6	1	3	1	1
6	0	300	4	2	3	2	2
7	400	200	3	4	5	2	2
8	100	200	6	2	3	2	2
9	100	0	2	3	4	2	2
10	300	500	4	4	4	2	2

Table 1: Data for the example instance.

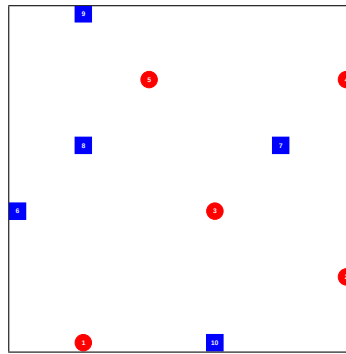


Figure 3: Instance with 2 types of items frequency to cover in a period of 3 days.

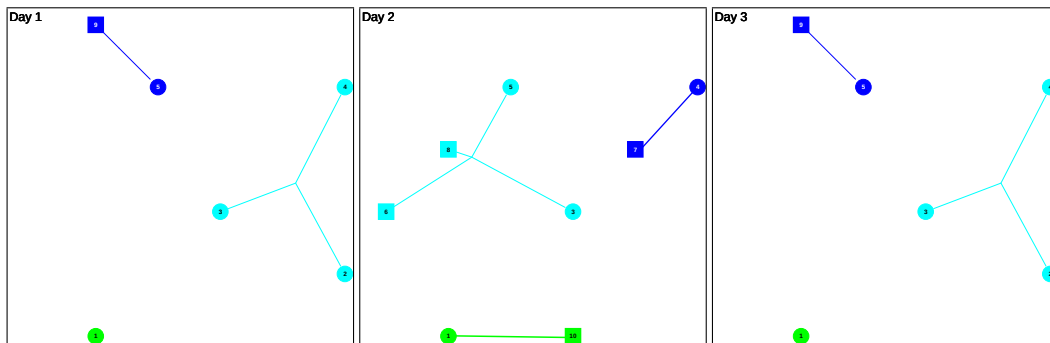


Figure 4: Solution visualization of the $gPCCCP$ instance with a period of 3 days.

3. Instance evaluation and results

Because of space requirements, we concentrate our tests to $gPHCCCP$ instances only. The other problem ($PFC-HCCCP'$) returns the Fermat-Weber points as the center of the clusters, and takes normally longer to be solved.

We developed our tests with five sets of instances: basic (cycle, 2cycle), pre-real (HOSPTEST) and real (HOSP1994). The set of basic ten instances were created with ($n = 5$ and 10 , $horizon = 3$), but a very disperse demand per day, and three different clusters' capacity and fixed costs. They were developed to be "toy" examples to obtain the preliminary view of the final topology of the formed clusters.

The second set of five instances called here as "pre-real" were developed from a subset of first 20 health units from the third set. The health units were covered in 6 days (week) with the UTM coordinates related to their spatial position in the city of Fortaleza/CE.

The third set with two instances we have 78 health units that were covered in 1994 every 6 days/week. We use their real UTM coordinates. The demands per day of the week are different for each day for any unit and the periodicity are classified accordingly the importance of the health unit. The clusters differ in capacity and fixed cost for both instances. These instances were extracted from a real life project developed for health waste collection, Negreiros [1994].

We selected five solvers to run our tests: Baron, Artelys Knitro, XPress, Gurobi and CPLEX. All instances were directly executed by using NEOS-Server from Argonne National Labs, Dolan [2001]. It was given to all solvers 10.8k-15ksec for each run after obtaining a new feasible solution, or the default time of the solvers if no feasible solution was found. In special cases we improved the time to 15ksec for some instances to obtain an upper bound for the instance.

We used AMPL as the programming language to develop our models and instances. The solvers were selected because they can run the basic gPHCCCP non-linear model (BARON and Artelys Knitro), or the second-order cone constraints (Artelys Knitro, Xpress, Gurobi and CPLEX) for the gPHCCCP.

In the following tables (2) and (3) the legends indicate: **Instance** the name of the instance, **Ind** is the index of the instance referred in the model, **NItems** is the number of items, **Period** is the number of days of a period, **SNCI** is the available number of clusters, (Q_j, FC_j) is the tuple referring the capacity and fixed cost of opening cluster j respectively.

In table (3) **NCI** means the solution, **Desc** means if the description of the feasible solution was obtained, **OF** means the objective function, **Time (s)** means the solve time in seconds reported by the AMPL platform. In the results, we have indicating the message reported by the solver: (*) Optimal solution found; (o) Optimal in a mingap; (v) Best feasible solution found; (Comp) a complete description of a solution was obtained; (x ISF) x integer solutions were found; (TL-NFS) Time limit with no solution found; (NO MEMORY) No memory to run the model; (None) No description provided or returned; (Error) Error in evaluating the instance.

Instance	IND	Basic Parameters			
		NItems	Period	SNCI	(Q_j, FC_j)
Cycle01	1	5	3	3	(5,10), (7,20), (8,30)
Cycle02	2	5	3	3	
Cycle03	3	5	3	3	
Cycle04	4	5	3	3	
Cycle05	5	5	3	3	

Table 2: Characteristics of the instances class Cycle.

In the instances of the class Cycle, we obtained complete results from both solvers Baron and Artelys Knitro, with an advantage in solution optimality obtained by Artelys Knitro, using the basic model, Table 3. It is notorious the "optimality proof" given by Baron, that is inadequate.

The results using the SQOP model for the class Cycle are reported in Table 4. It can be

Instance	IND	BARON (21.1.13)				Artelys Knitro (13.0.1)			
		NCI	Desc	OF	Time(s)	NCI	Desc	OF	Time(s)
Cycle01	1	3*	Comp	420060	0.132	3*	Comp	283393.3333	710.988
Cycle02	2	2*	Comp	140040	499.875	3*	Comp	60.0000	4581.110
Cycle03	3	2*	Comp	290050	500.680	3*	Comp	25060.0000	7509.180
Cycle04	4	2*	Comp	315050	499.693	3*	Comp	70060.0000	2842.660
Cycle05	5	1*	Comp	200030	498.877	3*	Comp	60.0000	3421.680

Table 3: Results of the instances class Cycle solved by the basic model (Baron) and SOCP model (Artelys Knitro).

seen the results are the same for most runs with minor difference, and the close dominance in time to obtain the solutions for XPress and Gurobi.

Instance	IND	XPRESS (v 8.13.1)				Gurobi (9.1.2)				CPlex (20.1.0.0)			
		NCI	Desc	OF	Time (s)	NCI	Desc	OF	Time (s)	NCI	Desc	OF	Time (s)
Cycle-01	1	3*	Comp	233393.32	93.7	3*	Comp	233393.33	16.1	3*	Comp	233393.54	19.9
Cycle-02	2	3o	Comp	60.00	0.8	3*	Comp	60.00	1.5	3*	Comp	60.00	711.1
Cycle-03	3	3o	Comp	250599.99	17.3	3*	Comp	25060.00	10.7	3o	Comp	25060.04	271.8
Cycle-04	4	3o	Comp	70059.99	25.3	3*	Comp	70059.99	19.8	3o	Comp	70060.19	183.9
Cycle-05	5	3*	Comp	60.00	2.4	3*	Comp	60.00	2.5	3o	Comp	60.00	180.0

Table 4: Results of the instances class Cycle solved by the SQOP model using XPress, Gurobi and CPLEX.

Instance	IND	Basic Parameters			
		NItems	Period	SNCI	(Q_j, FC_j)
2Cycle01	11	10	3	3	(10, 10k), (12, 20k), (15, 40k)
2Cycle02	12	10	3	3	
2Cycle03	13	10	3	3	
2Cycle04	14	10	3	3	
2Cycle05	15	10	3	3	

Table 5: Characteristics of the instances class 2Cycle.

Instance	IND	XPRESS (v 8.13.1)				Gurobi (9.1.2)				CPlex (20.1.0.0)			
		NCI	Desc	OF	Time (s)	NCI	Desc	OF	Time (s)	NCI	Desc	OF	Time (s)
2Cycle01	11	3o	Comp	834166.65	383.0	3*	Comp	834166.66	129.7	3*	Comp	834170.41	380.3
2Cycle02	12	2o	Comp	401666.66	9121.3	2*	Comp	401666.66	4516.2	2v	Comp	401668.27	29448.9
2Cycle03	13	3o	Comp	413333.33	39477.9	3v	Comp	442776.72	38984.7	3v	Comp	388333.33	43874.9
2Cycle04	14	2o	Comp	444000.01	656.4	3v	Comp	126666.64	43254.5	2*	Comp	444001.14	5599.8
2Cycle05	15	3o	Comp	126666.66	36.2	3*	Comp	126666.66	39076.0	3v	Comp	130000.73	30236.8

Table 6: Results of the instances class 2Cycle solved by the SQOP model using XPress, Gurobi and CPLEX.

In table 6 we have the resolution of the instances 2Cycle with 10 items and 3 days on a period (Table 1: instchar2C) using SQOP model or (gPCCCP), by the Quadratic MIP solvers XPress, Gurobi and CPLEX with Neos-Server. We see that the solvers did not prove optimality for major instances, but only for 3 of them (11, 12 and 15). Gurobi obtained the better average solutions although CPLEX returned one solution better than the others. It can be noticed that CPLEX proved optimal the instance 2Cycle04. Although agreeing with XPress, Gurobi found a better solution (152.38% better) and did not prove its optimality during the time given to finish the process.

The set of instances HOSP arose from a real life application performed in 1994 by Negreiros [1994] to a waste collection company in Fortaleza/CE. The first set, HOSPTEST, is for preliminary evaluation and the class HOSP1994 are the real instances used to calculate the periodic routes of the health waste collection service. The final solution reported by Negreiros [1994] had

4 trips (4 clusters of 8-8.8k tons) and it is related to the instance HOSP1994-01. The best result here provided, Table 8 use 6 clusters, as obtained with GUROBI. But if we adjust α from $10e^{-2}$ in instance HOSP1994-01 to $10e^{-5}$ in the same instance, the variable cost reduce the impact and the solution goes to 4 clusters, see instance HOSP1994-03 in Table 8.

In Table 8 it can be seen for instances HOSPTEST and HOSP1994. For HOSPTEST instance Cplex achieve notorious dominance over the other solvers, considering the final objective value. In figure 5 we show the solution obtained with XPress, for instance HOSPTEST-01. In fact it can be considered that is not a fair result once it was given more time to Cplex than to the others. Anyway, looking for the first instance it can be seen that the time given to XPress was longer, and the upper bound obtained was still far from the one obtained by Cplex because of its cutting methods that behaves better for these instances. For the instances HOSP1994 the absolute dominance in the objective value as for the time spent goes to Gurobi. These instances are almost 3.5 times bigger than previous set, and Gurobi return feasible solutions with impressive upper bound over XPress and Cplex. Important to notice that XPress return both solutions from the “root” indicating optimality, that is inappropriate.

Instance	IND	Basic Parameters			
		NItems	Period	SNCI	(Q_j, FC_j)
HOSPTEST-01	21	20	6	3	(8k,1k) (8.8k,1.1k) (10k,1.5k)
HOSPTEST-02	22	20	6	3	
HOSPTEST-03	23	20	6	3	
HOSPTEST-04	24	20	6	3	
HOSPTEST-05	25	20	6	3	
HOSP1994-01	31	78	6	6	$\alpha = 1.10^{-2}$
HOSP1994-02	32	78	6	6	$2x(8k,10k) 2x(8.8k,12.5k) (10k,18k) (11k,20k)$
HOSP1994-03	311	78	6	6	$\alpha = 1.10^{-5}$
HOSP1994-04	321	78	6	6	

Table 7: Characteristics of the instances class HOSPTEST and HOSP1994.

Instance	IND	XPRESS (v 8.13.1)				Gurobi (9.1.2)				Cplex (20.1.0.0)			
		NCI	Desc	OF	Time (s)	NCI	Desc	OF	Time (s)	NCI	Desc	OF	Time (s)
HOSPTEST-01	21	2v	17 ISF	935607136.24	67983.8			NO MEMORY		3v	Comp	458873637.92	43823.7
HOSPTEST-02	22	1v	8 ISF	1437039166.68	379.8	1v	Comp	1448032244.49	17.1	3v	Comp	473331254.17	31009.3
HOSPTEST-03	23	3v	2 ISF	1262676744.76	8638.1	1*	Comp	1161567846.35	48049.6	3v	Comp	459943047.41	43799.3
HOSPTEST-04	24	2v	2 ISF	1057669435.30	7374.9	1v	Comp	789931737.57	7375.6	3v	Comp	197738515.68	43869.7
HOSPTEST-05	25	2v	2 ISF	1258839757.64	3566.0	1v	Comp	995777514.73	92.1	3v	Comp	3087575.00	29839.4
HOSP1994-01	31	4v	10 ISF	552875.2227	90348.1			NO MEMORY		6v	Comp	798995.0757	110363
HOSP1994-02	32			NO MEMORY		2v	Comp	578258.1072	96155.1	3v	Comp	782080.6872	109338
HOSP1994-03	311	3o	3 ISF	32501.0550	55.1162	3o	Comp	32500.8533	417.38	3o	Comp	32500.9920	111661
HOSP1994-04	321			NO MEMORY		2o	Comp	20000.5537	15878.7	3v	Comp	32500.5274	111260

Table 8: Results of the instances class HOSP solved by the SQOP model using XPress, Gurobi and Cplex.

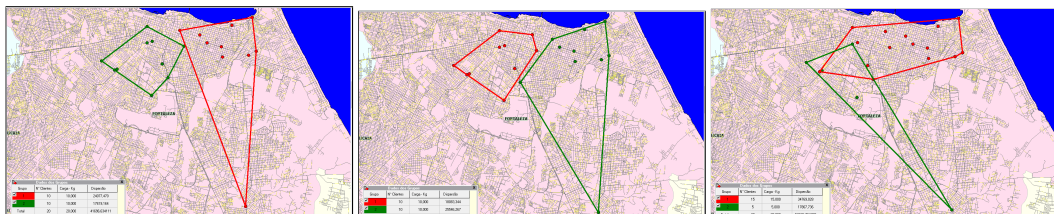


Figure 5: Xpress solution visualization of the HOSPTEST-01 instance with a period of 3 days.

4. Conclusion

Motivated by a problem that arises in the health waste collection and other applications, we proposed new models for two different versions of the Periodic Heterogeneous Capacitated Centred Clustering problem. In these versions, the Geometric Centers are searched to form clusters' centers and their assigned items by minimizing total internal variance and fixed cost to open clusters.

Because of its non-linear and non-convex nature, the proposed formulations is reduced to a quadratic-binary-mixed programming problem, where we show it is possible to be solved by sequential-quadratic using state-of-the-art MILP solvers (Xpress, Gurobi and CPLEX) for moderate sized instances (up to 78 items, six days of time horizon and 8 clusters). We also introduced a new set of constraints adjusted to deal with a continuous form to solve items' periodicity considering "time between visits", as a dual form of the established days of visits commonly used by the PVRP formulations.

References

- Araujo, E., Poldi, K., e Chaves, A. (2013). Clustering search applied to the periodic vehicle routing problem: A case study in waste collection. *Annals of XLV SBPO - Natal/RN*, p. 1916–1926.
- Baldacci, R., Bartolini, E., Mingozzi, A., e Valletta, A. (2011). An exact algorithm for the period routing problem. *Operations Research*, 59:228–241.
- Bard, J. e Jarrah, A. (2013). Integrating commercial and residential pickup and delivery networks: a case study. *Omega*, 14:703–720.
- Beltrami, E. e Bodin, L. (1974). Networks and vehicle routing for municipal waste collection. *Networks*, 4:65–94.
- Bianchi-Aguiar, T., Carravilla, M. A., e Oliveira, J. F. (2012). Municipal waste collection in ponte de lima, portugal – a vehicle routing application. *OR Insight*, 25:185–198.
- Boyd, S. e Vandenberghe, L. (2004). *Convex Optimization*. ISBN 978-0-521-83378-3.
- Campbell, A. e Wilson, J. (2014). Forty years of periodic vehicle routing. *Networks*, 63:22–15.
- Chartier, Y., Emmanuel, J., Pieper, U., Prüss, A., Rushbrook, P., Stringer, R., Townend, W., Wilburn, S., e Zghondi, R. (2014). *Safe management of wastes from health-care activities*. ISBN ISBN: 978 92 4 154856 4.
- Cordeau, F., Gendreau, M., e Laporte, G. (1997). A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks*, 30:105–119.
- Czyzyk, J., Mesnier, M., e Moré, J. (1998). The neos server. *IEEE Journal on Computational Science and Engineering*, 5(3):68–75.
- Dolan, E. (2001). The neos server 4.0 administrative guide. Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division, Argonne National Laboratory.
- Fawzi, H. (2019). On representing the positive semidefinite cone using the second-order cone. *Mathematical Programming*, 175:109–118. ISSN 0025-5610 ORCID 119324071.
- Fisher, M., Greenfield, A., Jaikumar, R., e Lester, J. (1982). A computerised vehicle routing application. *Interfaces*, 12:42–52.

- Fisher, M. e Jaikumar, R. (1981). A generalized assignment heuristic for vehicle routing. *Networks*, 11:109–124.
- Foster, B. e Ryan, D. (1976). An integer programming approach to the vehicle scheduling problem. *Oper Res Quart*, 27:367–384.
- Francis, P. e K., S. (2006). Modeling techniques for the periodic vehicle routing problems. *Transportation Research Part B*, 40:872–884.
- Gaudioso, M. e Palleta, G. (1992). A heuristic for the periodic vehicle routing problem. *Transportation Science*, 26:86–92.
- Gropp, W. e More, J. (1997). Optimization environments and the neos server. In Buhman, M. D. e Iserles, A., editors, *Approximation Theory and Optimization*, p. 167 – 182. Cambridge University Press.
- Huerta-Munoz, D., Archetti, C., Fernandez, E., e Perea, F. (2022). The heterogeneous flexible periodic vehicle routing problem: Mathematical formulations and solution algorithms. *Computers & Operations Research*, 141.
- Kalcsics, J. e Rios-Marcado, R. (2019). Districting problems. In Gilbert Laporte, F. S. d. G., Stefan Nickel, editor, *Location Sciences*, volume 284, p. 705–743. Springer International Publishing.
- Linhares, M., Pinto, R., e Maculan, N. (2022). New optimization models for clustering problems. *Submitted to EJOR*.
- Lobo, M., Vandenberghe, L., Boyd, S., e Lebret, H. (2019). Applications of second-order cone programming. *Linear Algebra and Its Applications*, 284:193–228.
- Negreiros, M. (1994). Avaliação de desempenho do sistema sisrot na coleta hospitalar. Relatório final, Consoft Consultoria Ltda. – Marquise Construções.
- Negreiros, M., A.F., X., Xavier, A., Maculan, N., Michelon, P., Lima, J., e Andrade, L. (2011). Optimization models, statistical and dss tools for prevention and combat of dengue disease. *IntechOpen*, Efficient Decision Support Systems, Chapter 7:115–160.
- Negreiros, M., Maculan, N., Batista, P., Rodrigues, A., e Palhano, A. (2020). Capacitated clustering problems applied to the layout of it-teams in software factories. *Annals of Operations Research*.
- Negreiros, M., Maculan, N., Palhano, A., Muritiba, A., e Batista, P. (2022). *Capacitated clustering models to real life applications*. ISBN 978-1-80355-982-7.
- Negreiros, M. e Palhano, A. (2006). The capacitated centred clustering problem. *Computers & Operations Research*, 33:1639–1663.
- Rios-Marcado, R. (2020). Research trends in optimization of districting systems. In Rios-Marcado, R. Z., editor, *Optimal Districting and Territory Design*, volume 284, p. 3–8. Springer.
- Tan, C. e Beasley, J. (1984). A heuristic algorithm for the period vehicle routing problem. *OMEGA*, 12:497–504.

Taslimi, M., Batta, R., e Kwon, C. (2020). Medical waste collection considering transportation and storage risk. *Computers & Operations Research*, 120.

UNDPGEF, O. (2023). *MODULE 14: Off-site Transport and Storage of Healthcare Waste*. URL <https://cdn.who.int/media/docs/default-source/wash-documents/wash-in-hcf/training-modules-in-health-care-waste-management>.

Yanik, S., Kalcsics, J., Nickel, S., e Bozcaya, B. (2019). A multi-period multi-criteria districting problem applied to primary care scheme with gradual assignment. *ITOR*, 26:1676–1697.