RCMeans: A Recursive Capacitated Means for Districting Problem

Polyana Bezerra da Costa[1], Italo Francyles Santos da Silva[1], Pedro Henrique Carvalho Vieira[1], Robert Douglas de Araújo Santos[1], Mayara Gomes Silva[1], Christyellen de Souza Costa Lima[1], Daniel Lima Gomes Júnior[1,2], Eliana Márcia Garros[3], Ítalo Fernandes Serra da Silva[3], e Lucas de Paula Assunção Pinheiro[3]

[1]*Universidade Federal do Maranhão, Núcleo de Computação Aplicada, São Luís-MA, Brasil*
[2]*Instituto Federal do Maranhão, Departamento de Computação, São Luís-MA, Brasil*
[3]*Equatorial Energia, Companhia Energética do Maranhão, São Luís-MA, Brasil*

Abstract—The billing process of an energy distributor in Brazil is connected to reading energy consumption logistics. An efficient, balanced and capacity enabled process has benefits with cost reducing and quality perception of the service provided. In capacitated clustering, the elements are associated with weights for construction of groups with limited capacities. This paper presents an approach to the capacitated clustering problem, applied to the consumer unit measurement groups organization in Brazil's energy distributors companies. The process of creating those measurement groups, in general, is carried out manually by expert analysts. The purpose of this problem modality is to create partitions that minimize the internal dispersion of the associated group. In this work, the RCMeans method, which is based on the K-Means technique applied to data groupings with the inclusion of the capacity constraint for the group definition, is presented. The obtained results show a comparison between the current situation and the result with the proposed method, under the cohesion analysis, separation, number of groups, silhouette index, and consumer units measurement mean time of the groups.

Index Terms—Capacitated Clustering, Reading Units, Electric Energy, Dispersion.

I. INTRODUCTION

The consumer unit measurement and invoice delivery for each end consumer are the main activities related to the billing process of an electrical energy distribution company. There are models of meter devices on the market that allow the telemetry solution [1] without needing manual measurement (realized by companies employees). However, in a big country such as Brazil, an effective change for these telemetry devices means replacing more than 63 million conventional meters [2] . So, we can infer that the manual measurement process of electricity consumption is and will be a problem for a reasonable time.

In Brazil, the conditions for supplying electrical energy are regulated by the National Electric Energy Agency (ANEEL). This regulation aims to improve the relationship between the providers of the public electricity and end consumers. Among the regulations, ANEEL establishes that a concessionaire has certain intervals to finish the measurement process [3]. This type of obligation hinders the management of the Reading Units (UL) in the measurements arranging and invoices deliveries tasks.

The energy company of Maranhão (CEMAR) and power plants from Pará S.A (CELPA) organize the measurement task and delivery of the invoices by creating groups of end customers, which can be understood as clusters. Thus, each measurement employee must have routes to perform the measuring and delivering invoices in each group. In practice, each municipality or region has an individual organization of measurement groups and subgroups.

The homogeneity and compactness criteria are used to size definition of these groups and selection of the consumer units. Homogeneity refers to the creation of groups that should be as homogeneous as possible, regarding the workload of the measurement employees, leading to the labor costs minimization. On the other hand, the compactness refers to the geographical form of the groups, which should be as compact as possible. This second feature contributes to efficient route generation. Moreover, other company activities as commercial management can use this group definition criterion, reducing the employee measurement total time, invoice delivery, power cuts, power reclose and fraud inspection.

The creation of these groups is defined as a Districting Problem (DP). The objective of the districting problem is to group small territorial units into contiguous and non-overlapping districts, given a certain objective function and a set of constraints.

The classic application of this problem is related to the definition of electoral districts [4], [5], [6]. In electoral districts problem, a geographical area should be divided into constituencies, each represented by a member of parliament. A wide number of applications can be enumerated, such as the public transport network [7], the definition of sales areas or seller workplaces [8], [9], [10], [11] and definition of areas for manufacturers and consumers [12].

Some methods that have been proposed to solve the districting problem are tabu search [13], evolutionary algorithm [14], simulated annealing, old bachelor acceptance algorithm [15] and GRASP [16].

The capacity grouping/districting problem (CDP) is a

generalization of the districting problem where districts have limited capacity.

These aspects highlight the importance of developing computational methods that are capable of efficiently handling the logistics of measurement groups, enabling the improvement of billing process by resources optimizing such as the energy consumption of consumer units measurement activity and delivering invoices. In addition, it can provide a consequent improvement in the quality of service with the index of customer satisfaction growth.

This paper presents the Recursive Capacitated Means (RCMeans), a method for creating homogeneous and compact groups of the consumer units, applied to the electric energy distribution companies context in Brazil. In addition, it presents a comparative analysis between the grouping performed manually by the company experts with the proposed RCMeans result.

The result of this work is part of a Research and Development (R&D) project, hired by CEMAR / CELPA (ANEEL PD-0371-0031/2017), executed by the Applied Computing Center (NCA) from the Federal University of Maranhão (UFMA).

II. Capacitated Clustering Problem

In capacitated clustering, a set of elements is subdivided in a set of clusters. Each element has a weight and it must belong to an only one cluster. The elements of a cluster may be, for instance, the consumer units of a particular city.

Each cluster has a predefined capacity. The total of weights associated to each cluster's element may not be greater than the capacity previously defined.

The capacitated clustering problems must satisfy the following requirements:

- one weight is associated with each element;
- each element must be associated with an only one cluster;
- the elements must be divided into p fixed groups;
- all the elements must belong to a group;
- the sum of weights for each element of a group must not be greater than the previously defined capacity;
- a criterion to determine the proximity/distance between grouped elements is required;

In clustering problems, it is necessary to determine a criterion to measure the similarity or dissimilarity between the elements.

A. Mathematical Model

According to [17], considering $O = \{1, 2, ..., n\}$ a set of elements to be grouped, we assume D_{kl} is a matrix n x n, formed by the distances of elements of O, where D_{kl} is the Euclidian distance between k and l, $D_{kl} > 0$, $D_{kl}D_{lk}$, $D_{kk} = 0$. Each element has a weight d_k, with $d_k > 0$.

Let $P = \{A1, A1, ..., Ap\}$ be the division of O in p groups. In addition, let $Q = \{Q1, Q2, ..., Qp\}$ be the set of medians or centers of P. The capacitated grouping problem proposed by [17] is presented in equation 1 (objective function), subjected to equations 2, 3 and 4.

$$f(P) = Min \sum_{k \in O} \sum_{Q_i \in Q, i \in I} D_{kQ_i} x_{kQ_i} \quad (1)$$

Equation 1 shows the objective function which minimizes the dispersion of the elements in each cluster. Equation 2 grants that when dividing O in p groups, p medians will be given. Equation 3 grants that each element will be assigned to a single median Q_i. Lastly, equation 4 specifies that the sum of the weights of each element assigned to group i cannot exceed the maximum capacity.

$$\sum_{k \in O} y_k = p \quad (2)$$

$$\sum_{Q_i \in Q, i \in I} x_{kQ_i} = 1, in\ which\ k \in 0, \forall k \quad (3)$$

$$\sum_{k \in O} d_k x_{xQ_i} \le C_i, in\ which\ Q_i \in Q, i \in I, \forall i \quad (4)$$

Where:

$$x_{kQ_i} = \begin{cases} 1, if\ the\ k^{th}\ element\ is\ assigned\ to\ the\ median \\ 0, if\ otherwise \end{cases}$$

$$y_k = \{1, if\ the\ k^{th}\ element\ is\ the\ median$$

III. Capacitated Clustering Problem

This section introduces a method in order to solve the capacitated clustering problem applied in the power companies' scenario. Figure 1 shows the four steps of the proposed method.
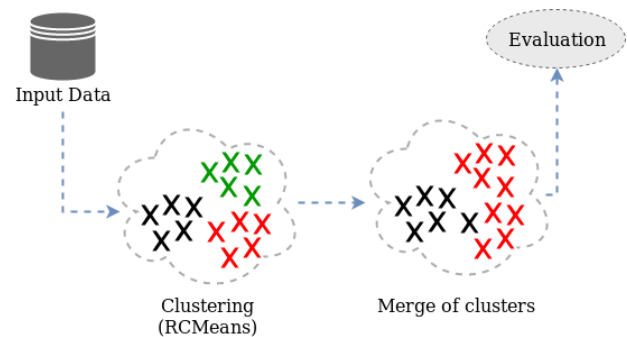

Fig. 1. Steps of the proposed method.

Firstly, Latitude and Longitude coordinates and the time of measuring of each consumer unit are collected from the dataset. This two information form a point. The data is then used as input for the RCMeans clustering algorithm. In this step, many groups are created. However, some of them contain a small number of points. Hence, these generated clusters are submitted to the merging step, where little clusters are merged based on a proximity criterion. The last step is the evaluation of the generated groups by applying clustering evaluation metrics.

A. RCMeans Algorithm

The RCMeans algorithm is proposed in this work. It is a method deployed to solve the capacitated clustering problem satisfying that constraints showed in Section II.

The RCMeans algorithm belongs to the unsupervised learn- ing category. It is based on the K-Means technique, which is quite used to perform the clustering task. In this one, the data of a set are splitted into groups according with a similarity criterion. The capacitated version of this algorithm, which is used in this work, also includes a capacity constraint for group generation. Figure 2 shows how the RCMeans algorithm works, and algorithm 1 presents an overview of the method.
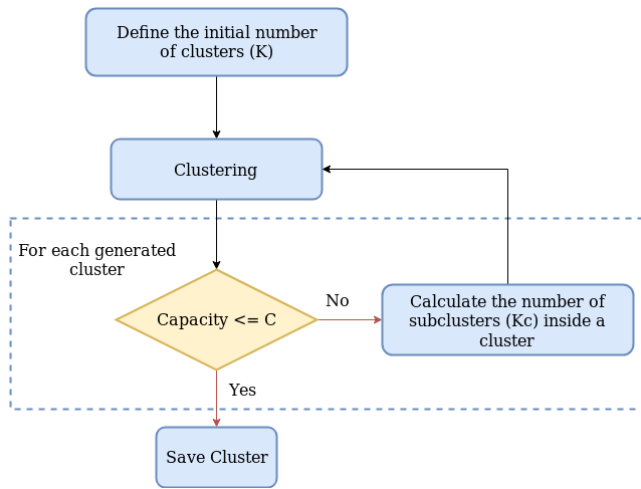
Fig. 2. Workflow of RCMeans.

The initial number of clusters (K) is based on the total of points and on the established capacity. An initial clustering is performed based on the number K. The capacity of each created group is evaluated and, if it exceeds the maximum capacity added to a tolerance value, the points of a group will be separated for a recursive clustering process. The new generated groups will also be evaluated. The steps of re-clustering, evaluation and group split are repeated recursively until all the generated groups satisfy the constraints. It is important to note that clusters whose capacity satisfies the constraint are preserved.

There are two guarantees provided by this algorithm: (1) all points belong to a single cluster, and (2) all groups satisfy the desired capacity. To avoid maintaining groups with small capacities, a merge of adjacent clusters is performed in order to ensure the creation of more homogeneous groups.
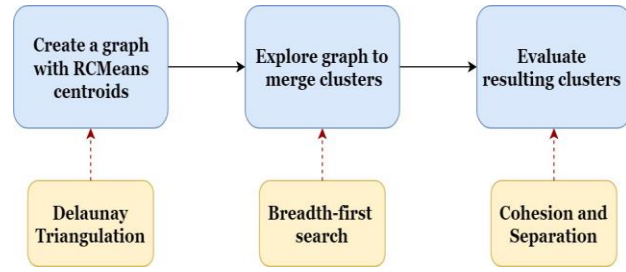
B. Merging Clusters

Fig. 3. Steps in the process of merging clusters.

The RCMeans algorithm produces groups that will always respect the maximum capacity, although some clusters may have a very moderate capacity. A post processing step is necessary to merge small clusters in order to produce groups with fuller capacity. An under capacity cluster can be merged with other clusters until they reach the maximum capacity, which is 360 minutes per cluster (daily productivity time of each measuring employee).

Algorithm 1 RCMeans

input : *points*: Coordinates of every unit in the city and its average reading time.
output: Set of clusters.

    1. Define C as the maximum capacity of each cluster;
  2. Define the initial number K of centroids;
  3. Choose K random centroids;

procedure RCMEANS(*points, K*)
while *centroids change* **do**
   for *each point in points* **do**
     for *each centroid in* $K_{centroids}$ **do**
       4. Calculate the Euclidian distance between p and c;
     end
     5. Include p in the cluster with the closest centroid to p;
   end
   6. Update the centroid of each cluster;
end
for *each cluster in Clusters* **do**
   7. $c_{cap} \leftarrow$ cluster's capacity;
   if $c_{cap} \leq C$ **then**
     8. Save cluster;
   else
     9. $M \leftarrow$ mean time of each cluster;
     10. $Np \leftarrow$ number of points in the *cluster;*
     11. $i \leftarrow C_{cap/M}$
     12. $K \leftarrow Np/i$
     13. RCMEANS(points, $K_{c)}$;
end
end procedure

Besides the capacity, a small cluster must be merged with a closer cluster, otherwise the compactness of the group will decrease. In order to merge clusters based on their capacity

and proximity, a graph that connects them is created. Each cluster on the graph will be a vertex represented by the cluster's centroid; the edges represent the connection between two centroids, and the weights of the edges are the Euclidean distance between the points. The approach used to construct the graph and connect the vertices was based on Delaunay Triangulation.

Building a complete graph (a graph with an edge connecting each pair of vertices) would be impractical due to the running time necessary to calculate the weights of each edge, which are the Euclidean distances between two centroids. Furthermore, it is not desirable to connect each centroid to every other centroid in the graph because this could lead to the merging of clusters that are not closely related. Therefore, a non-complete graph would suffice. Moreover, [18] remarks that there is a planar graph almost as good as a complete graph, which could be achieved with a Delaunay Triangulation.

According to [19], the Delaunay Triangulation divides a set of N points ($N \geq 3$) within a plane into a collection of triangles, resulting in a planar graph. This method was chosen because of its computational efficiency; in its worst case, the running time for a large set of point is $O(N^2)$ [20]. On the other hand, this method can create edges between distant vertices. In order to eliminate those edges, a threshold based on the average length of the edges was set; if the edge is greater than the threshold, it is removed from the graph.

After building the graph, a Breadth-First searching algorithm (BFS) was used to search the graph in order to merge adjacent vertices. BFS is one of the most popular algorithms for searching through a graph and has time complexity of $O(V + E)$, where V is the number of vertices, and E is the number of Edges. As reported by [21], BFS discovers all vertices at distance k from a vertex before discovering any vertices at distance $k + 1$. Due to this, the closest neighbors of a vertex will be visited first and the proposed method will try and merge adjacent clusters. If the capacity of a vertex is not full, a BFS algorithm starting from that specific vertex will return the vertices reachable from that vertex. A vertex V will be merged with its neighbor(s) until the maximum capacity is reached. If the sum of the vertices' capacities is greater than the maximum, they cannot be merged. In this case, the difference between the maximum capacity and the capacity of vertex V will be calculated; let this difference be N. These N elements will be removed from an adjacent cluster C and added to the cluster represented by the vertex V, i.e, the N elements of cluster C that are closer to V. Each step of this process is shown in algorithm 2. This post processing step results in less clusters, with more elements in each cluster. The next step is the evaluation of the resulting clusters.

C. Cluster Evaluation

In order to evaluate the results of the proposed method, and to compare them with the current clustering situation defined manually by CEMAR and CELPA, validity measures were applied to both clustering data. According to [22], general cluster validity can be expressed in terms of cohesion and separation. Cohesion quantifies how close the objects within a cluster are, and it expresses the compactness of a group, while separation determines how isolated a cluster is from other clusters.

The overall cluster validity for a set of K clusters can be seen as a weighted sum of the validity functions of each cluster. The functions used to calculate cluster validity are expressed in equations 5 and 6. The proximity function is a two dimensional Euclidean distance. C_i represents the ith cluster, while x represents each object within C_i, and c_i is the centroid of the group; w_i is the weight of the validity function, in this case, each function was multiplied by 1. C_j is the closest cluster to C_i, while c_j is its centroid.

$$cohesion(C_i) = \sum_{x \in C_i} proximity(x, c_i) \quad (5)$$

$$separation(C_i, C_j) = proximity(c_i, c_j) \quad (6)$$

There is also a cluster evaluation measure called the Silhouette Coefficient (SC) that was used to evaluate each cluster individually. To calculate the SC of a certain cluster, a random i object of this cluster is chosen. The average distance from i to all the other objects in the cluster is then calculated, let this value be a_i. The average distance from i to all the objects of the closest cluster that does not contain i is calculated; this value is called b_i. Hence, the silhouette coefficient can be expressed as:

$$SC = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (7)$$

Algorithm 2 Merging Clusters

---

input : *clusters*: Resulting clusters from RCMeans and their centroids;
C : maximum capacity for each cluster.
output: Set of clusters.

1. $Cap \leftarrow$ the capacities of each cluster;
2. Get a planar graph of the set of centroids with Delaunay Triangulation;
3. Calculate the average size of an egde;
4. Cut edges that are greater than the average;
5. Build a graph *tt* with the centroids and the resulting edges from Delaunay Triangulation;
for *each vertex in tt* **do**
  6. $v_{cap} \leftarrow$ capacity of *vertex*;
  if $v_{cap} < C$ *and vertex not visited* **then**
    7. *vertices* $\leftarrow$ result of the breadth-first search starting in *vertex*;
    $A \leftarrow$ the amount of elements required to complete *vertex* capacity;
    for *each v in vertices* **do**
      9. $v\_i_{cap} \leftarrow$ capacity of *v;*
      if $v_{cap} + v\_i_{cap} \leq C$ **then**
        10. Merge *v* with *vertex;*

11. Update v_{cap};
12. Mark *v* as visited;
    else
        13. Get *A* elements from *v* that are closer to *vertex*;
        14. Remove those elements from *v* and include them in *vertex*;
        15. Update v_{cap};
        end
    end
    end
end
16. Save remaining vertices as the final clusters.

The SC has a range of values that vary from [-1, 1]. If the coefficient has a negative value, it means that the cluster is sparse. The closer to 1 the coefficient is, the more compact a cluster is. In a good clustering, all groups should have a positive silhouette coefficient. In regards of cohesion, higher values are better, but when it comes to separation, lower values are more desirable.

IV. RESULTS AND DISCUSSION

The proposed method was tested in 2 different cities: one metropolis and one mildly populated city. QGIS [23] was used to plot results and to visualize geospatial information. Table I shows the results of the validity measures presented in section III-C. Each acronym mentioned on Table I has the following meaning:

- N : number of consumer units in the whole city
- UL: number of clusters (Reading Units)
- Coh.: overall clustering cohesion
- Sep.: overall clustering separation
- SC: average silhouette coefficient for all clusters
- MTs: mean time in seconds for reading each consumer unit

Table I presents the results comparing the current state with an improved stated generated by the proposed RCMeans. Current state was constructed over an manual optimization using general information.

The results shows that for São Luís, the RCMeans outperforms the current state used by the company, for all validity measures. São Luís has more urban and industrial land occupation. The number of clusters decreased, which shows that the new groups have fuller capacities. Consequently, the overall clustering cohesion and separation improved in comparison with the current clustering situation.

The silhouette coefficient presented a positive value, while the same coefficient for the current clustering situation resulted in a negative value. This means that the groups in the new clustering are more isolated from their neighbors, and elements of a same group are closer to each other, resulting in a more compact group. On the other hand, the clusters defined manually by CEMAR are sparser, which reduces the compactness of a cluster.

Figure 4 presents the overall result in São Luís, showing the current state in (a) and the resulting RCMeans clusters in (b). For better understanding, Figure 5 presents the downtown region of the city, where line shaped clusters happens in the manually defined districts in letter (a) which also ruins the overall separation since there can be parallel clusters (highlighted in Figure 6(a)). Furthermore, this is bad for the meter reader, because instead of exploring a compact area completely, the reader will have to walk through an extensive avenue in order to complete the required working hours. Differently, the RCMeans clustering results seen in Figure 5(b) show the same region with circular or more compact clusters, highlighted in Figure 6(b). Also, RCmeans has led to a drastic reduction of RUs required to perform the same activity while promoting fairness of work effort for all meter readers. The reduction found was approximately 35%.

For Santarém - PA, a city that has a large part of its area in rural or riverside regions, we observed that the cohesion of the groups obtained lower indices than those defined manually. This result is expected because the user experience was used in the manual definition to better define the groups for those regions of difficult access. Figure 7 shows a comparison between the clustering of the same region. The first region was defined manually and the second one by RCMeans. We observed that the user experience leads to the definition of groups with few RUs whereas our proposed method tends to merge small neighbor groups. Therefore, our merge strategy generated sparse groups for that region, and it decreased cohesion indices (this analysis is demonstrated in Figure 8). Even so, the silhouette coefficient between the groups shows a better result, demonstrating that the groups generated behave better and more efficiently than those determined manually. This is proven by the number of RUs generated approximately 7% lower. The lower number of RUs generates the allocation of less labor required to perform the same task, which directly impacts the operational costs of the activity.

Regarding the reading time for each consumer unit, Table 1 shows that the average time decreased 43% in São Luís, and just 2% in Santarém, mostly because of the clustering in rural areas.

V. CONCLUSION

The logistics process of reading and billing for electric consumption carried out by electric companies is a large and costly task. The organization of this process into reading units makes it possible to create job sequencing and reduce associated operating costs. In addition, automation and optimization make it possible to intelligently and quickly maintain information such as creating new customers or updating the status of old customers.

This paper proposes the RCMeans to solving the districting problem. Results show that the presented RCMeans method has generated better clusters in comparison with the actual clustering approach. For all executed evaluations, the silhouette is better when RCMeans is applied indicating a balancing coefficient over cohesion, separation, and compacity. We also get lower mean time for reading each consumer unit.

Another important result is the amount of generated groups. As long in the actual scenario, there are 1450 ULs, in RCMeans it was achieved a solution with only 938 clusters with the same amount of elements, which indicates for compa- nies that the same work realized nowadays can be performed with fewer measuring employees.

For future works, we intend to evaluate the technique over more cities and use other techniques for comparison purposes with the method proposed in this work. Some advances will be made to deal with difficult to reach areas such as constraining the geometry of clusters and thus generating more linear clusters or proposing differentiated methods to deal with these situations in order to create different specializations according to the situation.

REFERENCES

[1] S.-t. WANG, H. ZHOU, R.-m. YUAN, Z.-l. YI, Concept and application of smart meter [j], Power System Technology 4 (2010) 006.

[2] ANEEL, Aneel recebe sugestões sobre novo medidor inteligente de energia (sep 2010).
URL: http://www.brasil.gov.br/governo/2010/09/ aneel-abre-audiencia-publica-sobre-novo-modelo-de-medidor-de-energia

[3] ANEEL, Resolução normativa n° 414 (sep 2010).
URL: http://www.aneel.gov.br/documents/656877/14486448/ bren2010414.pdf/3bd33297-26f9-4ddf-94c3-f01d76d6f14a?version=1.0

[4] F. Ricca, A. Scozzari, B. Simeone, Political districting: from classical models to recent approaches, 4OR 9 (3) (2011) 223. doi:10.1007/s10288-011-0177-5.
URL: https://doi.org/10.1007/s10288-011-0177-5

[5] R. S. Garfinkel, G. L. Nemhauser, Optimal political districting by implicit enumeration techniques, Management Science 16 (8) (1970) B495–B508.

[6] M. Hojati, Optimal political districting, Computers & Operations Research 23 (12) (1996) 1147 – 1161. doi:https://doi.org/10.1016/S03050548(96)00029-9
URL:http://www.sciencedirect.com/science/article/pii/ S0305054896000299

[7] F. Tavares-Pereira, J. R. Figueira, V. Mousseau, B. Roy, Multiple criteria districting problems, Annals of Operations Research 154 (1) (2007) 69–92. doi:10.1007/s10479-007-0181-5.
URL: https://doi.org/10.1007/s10479-007-0181-5

[8] R. Z. Ríos-Mercado, E. Fernández, A reactive grasp for a commercial territory design problem with multiple balancing requirements, Comput. Oper. Res. 36 (3) (2009) 755–776. doi:10.1016/j.cor.2007.10.024.
URL: http://dx.doi.org/10.1016/j.cor.2007.10.024

[9] C. Easingwood, A heuristic approach to selecting sales regions and territories, Operational Research Quarterly (1970-1977) 24 (4) (1973) 527–534.d

[10] R. J. Shanker, R. E. Turner, A. A. Zoltners, Sales territory design: An integrated approach, Management Science 22 (3) (1975) 309–320. arXiv:https://doi.org/10.1287/mnsc.22.3.309, doi:10.1287/mnsc.22.3.309.

[11] S. W. Hess, S. A. Samuels, Experiences with a sales districting model: Criteria and implementation, Management Science 18 (4-part-ii) (1971) P–41–P–54. arXiv:https://doi.org/10.1287/mnsc.18.4.P41, doi:10.1287/mnsc.18.4.P41.
URL: https://doi.org/10.1287/mnsc.18.4.P41

[12] B. Fleischmann, J. N. Paraschis, Solving a large scale districting problem: a case report, Computers & Operations Research 15 (6) (1988) 521 – 533. doi:https://doi.org/10.1016/0305-0548(88)90048-2.
URL: http://www.sciencedirect.com/science/article/pii/0305054888900482

[13] B. Bozkaya, E. Erkut, G. Laporte, A tabu search heuristic and adaptive memory procedure for political districting, European Journal of Operational Research 144 (1) (2003) 12 – 26. doi:https://doi.org/10.1016/S0377-2217(01)00380-0.

[14] S. Ahmadi, I. H. Osman, Greedy random adaptive memory programming search for the capacitated clustering problem, European Journal of Operational Research 162 (1) (2005) 30 – 44, logistics: From Theory to Application. doi:https://doi.org/10.1016/j.ejor.2003.08.066.
URL: http://www.sciencedirect.com/science/article/pii/ S0377221703008208

[15] F. Ricca, B. Simeone, Local search algorithms for political districting, European Journal of Operational Research 189 (3) (2008) 1409 – 1426. doi:https://doi.org/10.1016/j.ejor.2006.08.065
URL: http://www.sciencedirect.com/science/article/pii/ S0377221707006273

[16] R. Z. Ríos-Mercado, J. C. Salazar-Acosta, A grasp with strategic oscillation for a commercial territory design problem with a routing budget constraint, in: I. Batyrshin, G. Sidorov (Eds.), Advances in Soft Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 307–318.

[17] J. M. Mulvey, M. P. Beck, Solving capacitated clustering problems, European Journal of Operational Research 18 (3) (1984) 339 – 348. doi:https://doi.org/10.1016/0377-2217(84)90155-3.
URL:dd http://www.sciencedirect.com/science/article/pii/ 0377221784901553

[18] P. Chew, There is a planar graph almost as good as the complete graph, in: Proceedings of the second annual symposium on Computational geometry, ACM, 1986, pp. 169–177.

[19] D.-T. Lee, B. J. Schachter, Two algorithms for constructing a delaunay triangulation, International Journal of Computer & Information Sciences 9 (3) (1980) 219–242.

[20] S. Sloan, A fast algorithm for constructing delaunay triangulations in the plane, Advances in Engineering Software (1978) 9 (1) (1987) 34–55.

[21] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to algorithms, MIT press, 2009.

[22] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Education India, Boston, MA, USA, 2007.

[23] Q. D. Team, et al., Qgis geographic information system, Open Source Geospatial Foundation Project. Available at em:< http://www. qgis. org/>.

TABLES AND FIGURES

TABLE I
RESULTS OF THE PROPOSED METHOD COMPARED WITH THE CURRENT DISTRIBUTION OF READING UNITS IN STUDIED POWER COMPANIES.

City	Method	N	ULs	Cohesion	Sep.	SC	TMs
São Luís (MA)	Current State	475.740	1.450	4.34	0.13	-0.056	54.13
	RCMeans	475.740	938	5.98	0.10	0.07	30.76
Santarém (PA)	Current State	103.356	488	23.9	1.87	-0.01	81.58
	RCMeans	103.356	457	5.1	0.21	0.14	80.11

(a)　　　　　　　　　　　　　　(b)

Fig. 4. Comparison results of (a) current clusters of São Luís - MA (Manually defined) with the RCMeans results of the same region in (b).
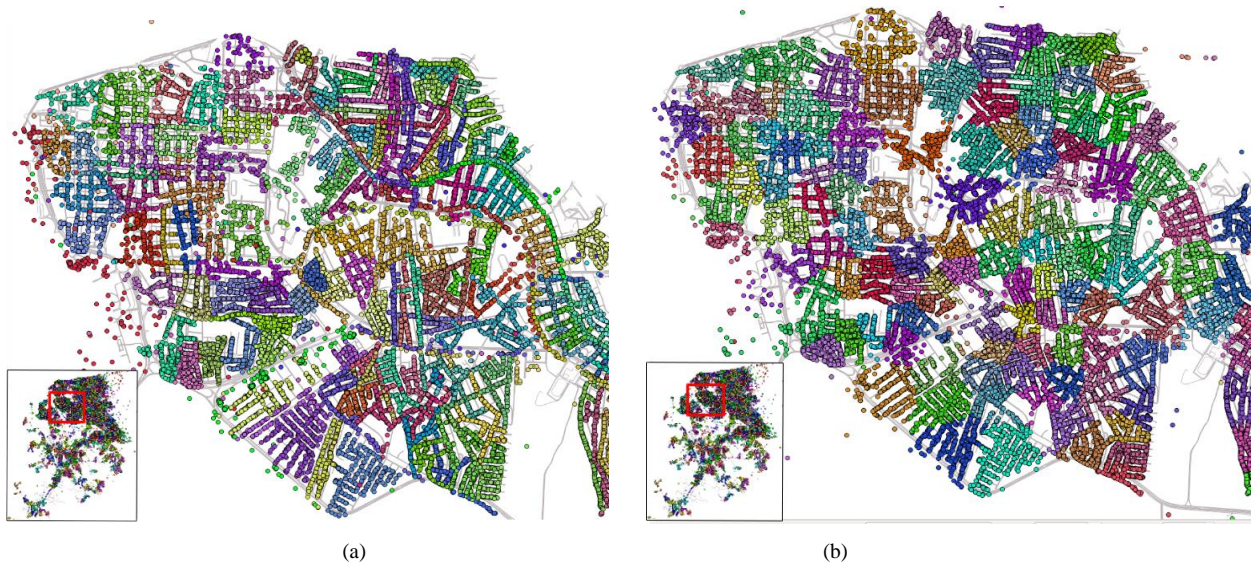

(a)　　　　　　　　　　　　　　(b)

Fig. 5. A closer look to a specific region of clusters in the current clustering situation. In (a), the current clustering and in (b) the improved version using RCMeans.
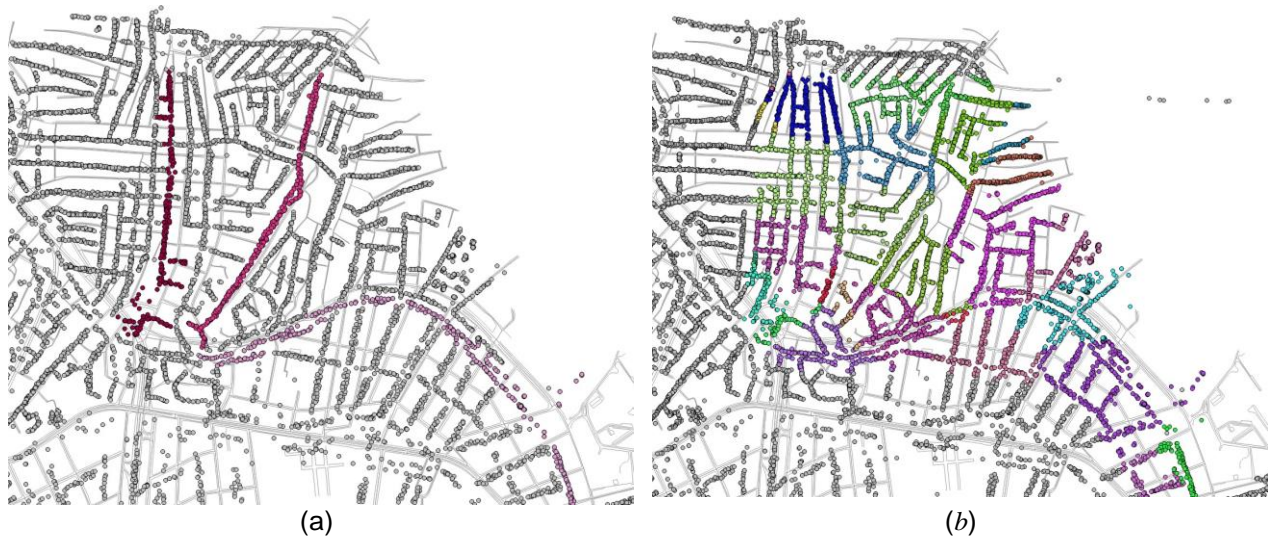
Fig. 6. Specific cluster generation in (b) with proposed RCMeans compared with line-shaped cluster generation of the current state in (a).
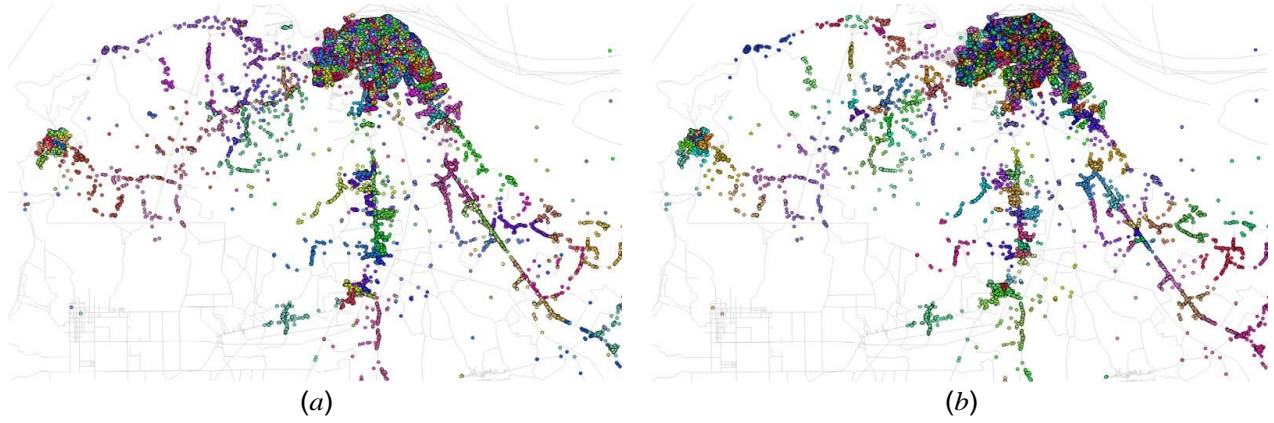

Fig. 7. Current clusters of Santarém - PA (Manually defined) in (a) and results of RCMeans clustering for Santarém - PA in (b).
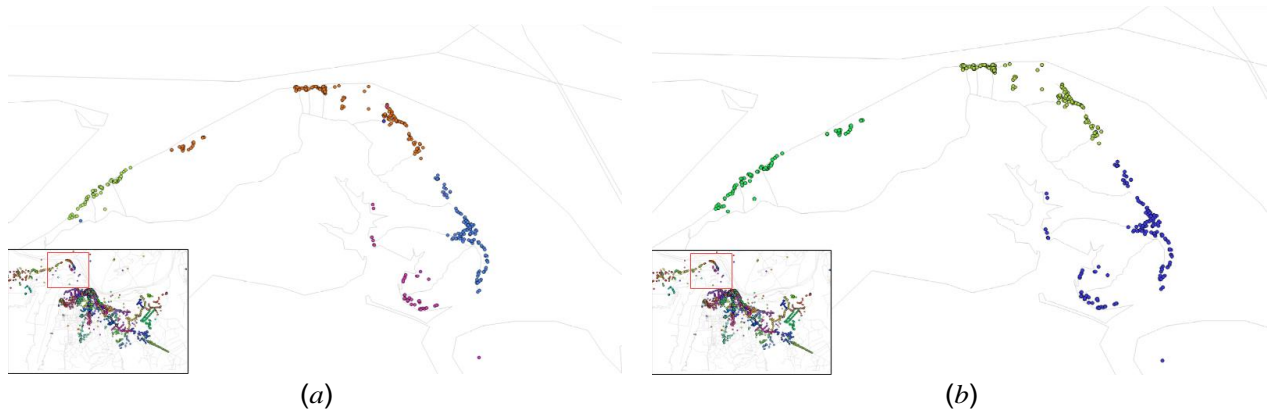

Fig. 8. View of a Satarém's rural regions. In (a) the current situation manually defined. In (b) the capacitated regions generated by RCMeans. We would notice a formation of sparse clusters in RCMeans result.