Netw Spat Econ (2011) 11:487–507
DOI 10.1007/s11067-010-9151-6

New Models for Commercial Territory Design

María Angélica Salazar-Aguilar ·
Roger Z. Ríos-Mercado ·
Mauricio Cabrera-Ríos

Published online: 7 January 2011

Abstract In this work, a series of novel formulations for a commercial territory design problem motivated by a real-world case are proposed. The problem consists on determining a partition of a set of units located in a territory that meets multiple criteria such as compactness, connectivity, and balance in terms of customers and product demand. Thus far, different versions of this problem have been approached with heuristics due to its NP-completeness. The proposed formulations are integer quadratic programming models that involve a smaller number of variables than heretofore required. These models have also enabled the development of an exact solution framework, the first ever derived for this problem, that is based on branch and bound and a cut generation strategy. The proposed method is empirically evaluated using several instances of the new quadratic models as well as of the existing linear models. The results show that the quadratic models allow solving larger instances than the linear counterparts. The former were also observed to require fewer iterations of the exact method to converge. Based on these results the combination of the

M. A. Salazar-Aguilar · R. Z. Ríos-Mercado (✉)
Graduate Program in Systems Engineering, Universidad Autónoma de Nuevo León,
San Nicolás de los Garza, NL, 66450, Mexico
e-mail: roger@yalma.fime.uanl.mx

M. A. Salazar-Aguilar
e-mail: angy@yalma.fime.uanl.mx

M. Cabrera-Ríos
Industrial Engineering Department, University of Puerto Rico at Mayaguez,
Mayaguez, PR, 00681-9000, Puerto Rico
e-mail: mauricio.cabrera1@upr.edu

quadratic formulation and the exact method are recommended to approach problem instances associated with medium-sized cities.

Keywords Mixed-integer linear programming · Integer quadratic programming · Territory design · Location · Valid inequalities

1 Introduction

Territory design or districting consists of dividing a set of basic units into subsets or groups according to specific planning criteria. In most applications, these basic units are city blocks, zip codes or individual customers and the resulting groups are known as territories or districts. A survey on general territory design problems (TDPs) can be found in Kalcsics et al. (2005). Two important applications of territory design are political districting (Bozkaya et al. 2003; Fleischmann and Paraschis 1988; Garfinkel and Nemhauser 1970; Hess et al. 1965; Hojati 1996; Mehrotra et al. 1998; Ricca and Simeone 2008; Shirabe 2009) and sales territory design (Drexl and Haase 1999; Hess and Samuels 1971; Marlin 1981; Zoltners and Sinha 2005). The characteristics of the problem addressed in this paper as detailed later, make it different to those studied in the work listed previously however.

In land use or site search problems, a set of compact territories is sought subject to connectivity constraints. The main difference with our problem is that the territories do not necessarily form a partition of the basic units. For recent models and methods on this type of problems, the reader is referred to the work of Aerts et al. (2003) and Xiao (2006).

The problem addressed in this work was motivated by the challenge faced by a local distribution company for bottled beverages where the objective was to create a specific number of territories given a set of city blocks (basic units). The territories were required to be compact, contiguous, and balanced in terms of number of customers and sales volume. Ríos-Mercado and Fernández (2009) introduced this problem initially with an initial solution approach based on a reactive GRASP (Greedy Randomized Adaptive Search Procedure), a metaheuristic procedure. Compactness in their initial work was modeled through the objective function of the p-center problem (pCP), which represents dispersion. Additionally, balanced territories in terms number of customers, sales volume, and workload were seeked out. With this set up, results were reported as better than those previously generated by the company hosting the study in terms of dispersion and balance requirements. Different versions of the problem have been studied by Segura-Ramiro et al. (2007) and Caballero-Hernández et al. (2007). In each of these, heuristic approaches were developed for large-sized instances that would be intractable for exact optimization purposes. Indeed, to the best of our knowledge, no exact scheme has been developed for neither of these models in the literature, only heuristic approaches can be found. Small and medium-sized instances, however, are also frequent enough in real life and therefore, their solutions are deemed important. The development of an exact optimization method that effectively handles the exponential number of connectivity constraints in small and medium-sized instances of the commercial territory design problem is, then, one of the key contributions of the work presented here.

On the other hand, in territory design problems, models dealing with connectivity constraints are usually approached through heuristics, as reviewed in Kalcsics et al. (2005), although a few works do provide optimal solutions, for example Garfinkel and Nemhauser (1970) and Shirabe (2009). The former studied a districting problem with 39 BUs and seven territories, while the latter proposed a solution method to a similar problem using 48 basic units to a variable number of territories. The method proposed in Shirabe (2009) was proved tractable only for a small number of territories.

Our work presents contribution in two directions regarding the commercial territory design problem. The first direction consists of an exact optimization procedure. The proposed algorithm is geared towards the solution of up to medium-sized instances of around 200 basic units to form up to around ten territories. The algorithm consists on the iterative solution of a mixed integer linear programming problem (MILP) through the relaxation of the connectivity constraints. The violated constraints are then identified through the solution of a simple separation problem. After that, these constraints are introduced as cuts to the model. The procedure continues until optimality is reached.

In the second direction, a new integer quadratic programming (IQP) formulation is proposed. This formulation greatly reduces the number of binary variables allowing the solution of larger instances than those allowed by the MILP counterpart. The exact optimization procedure is tested here with both, the MILP and the IQP formulations.

An empirical study on territory compactness over a wide range of instances is also presented to elucidate which kinds of measure have the potential to provide the best solutions for the commercial territory design problem. In general territory design, there is not a standard measure for compactness. One can find different kind of measures depending on the specific application. In the context of political districting, for instance, there have been some studies on compactness measures in Altman (1998). This criterion is discussed by Kalcsics et al. (2005) as well from a more general perspective, considering a median-based measure and a function based on convex hulls specifically tailored for their geometric approach. These works conclude that there is not a rigorous definition of this concept. In the absence of a standard measure for the case of commercial territory design, we carried out experimental work over a wide range of instances in order to analyze the performance of center- and median-based models.

The paper is organized as follows. Section 2 contains the description and mathematical formulations for this problem. Section 3 describes the proposed solution procedure. Experimental work is included in Section 4. Finally, conclusions are drawn in Section 5.

2 Problem description

Let $G = (V, E)$, be a graph where V is the set of basic units (BUs)—blocks in this case–and E is the set of edges representing adjacency between blocks. Each node j in set V has a series of parameters such as geographical coordinates (c_j^1, c_j^2), and two attributes or activities: number of customers and sales volume. An Euclidean distance, d_{ij} can be computed between each pair of BUs i and j. The set of BUs is to be partitioned into p territories, and it is required that each node is assigned to only one territory (exclusive assignment). The company seeks balanced territories with respect to the number of customers and product demand.

Let the size of territory $V_k \subset V$ with respect to activity a be defined as $w^{(a)}(V_k) = \sum_{i \in V_k}(w_i^{(a)})$, where k is a territory index, $a \in A = \{1, 2\}$ and $w_i^{(a)}$ is the value associated to activity a in node $i \in V$. Due to the discrete structure of the problem and to the unique assignment constraints, it is practically impossible to have perfectly balanced territories, i.e., territories of exactly the same size, with respect to each activity. Thus, in order to model balancing, a tolerance parameter $\tau^{(a)}$ for activity a is introduced. This parameter measures the relative deviation from the average territory size with respect to activity $a \in A$. This target average is given by $\mu^{(a)} = w^{(a)}(V)/p$. Another important constraint is connectivity, i.e., for each i and j assigned to the same territory there must exist in G a path between them totally contained in the territory. In addition, in order to pursue compactness, BUs of the same territory must be as close as possible to each other. One way to achieve this requirement is to minimize a dispersion measure. Several measures have been used in the literature. In this work we study two different measures, one based on the p-Center Problem (pCP) objective and the other based on the p-Median Problem (pMP) objective. This leads to two different models. Both are described next.

Formally, the problem consists in finding a p-partition of V according to the specified planning requirements of balancing and connectivity, that minimizes a given dispersion measure.

2.1 Mixed integer linear models

For the mathematical model MTDP (Median-Based TDP), the following set and decision variables are defined.

Set

N^i set of nodes adjacent to node i, where $N^i = \{j \in V : (i, j) \in E \vee (j, i) \in E\}, i \in V$.

Decision variables

$$x_{ij} = \begin{cases} 1 & \text{if basic unit } j \text{ is assigned to territory with center in } i;\ i, j \in V \\ 0 & \text{otherwise.} \end{cases}$$

Note that $x_{ii} = 1$ implies i is a territory center.

$$\text{(MTDP) minimize} z = \sum_{j \in V} \sum_{i \in V} d_{ij} x_{ij} \tag{1}$$

subject to:

$$\sum_{i \in V} x_{ii} = p \tag{2}$$

$$\sum_{i \in V} x_{ij} = 1 \qquad j \in V \tag{3}$$

$$\sum_{j \in V} w_j^{(a)} x_{ij} \geq (1 - \tau^{(a)}) \mu^{(a)} x_{ii} \quad i \in V; a \in A \tag{4}$$

$$\sum_{j \in V} w_j^{(a)} x_{ij} \leq (1 + \tau^{(a)}) \mu^{(a)} x_{ii} \quad i \in V; a \in A \tag{5}$$

$$\sum_{j \in \cup_{v \in S}(N^v \setminus S)} x_{ij} - \sum_{j \in S} x_{ij} \geq 1 - \mid S \mid \qquad i \in V, \quad S \subset [V \setminus (N^i \cup \{i\})] \tag{6}$$

$$x_{ij} \in \{0, 1\} \quad i, j \in V \tag{7}$$

Objective (1) represents a dispersion measure based on the pMP objective. In this sense minimizing dispersion is equivalent to maximizing compactness. Constraint (2) assures the creation of exactly p territories. Constraints (3) assure that each node is assigned to only one territory. Constraints (4) and (5) represent the territory balance with respect to each activity measure as they establish that the size of each territory must lie within a range (measured by tolerance parameter $\tau^{(a)}$) around its average size. Constraints (6) guarantee territory connectivity. They assure that for any given subset S of nodes assigned to center i not containing node i there must be an arc between S and the set containing i. They are similar to the subtour elimination constraints in the Traveling Salesman Problem. Note that there are an exponential number of such constraints so they cannot be explicitly written out. The proposed solution procedure generates only those that are needed in an iterative way. This model was used by Segura-Ramiro et al. (2007), and it can be viewed as a pMP problem with multiple capacity constraints, and with additional side constraints (4)–(6), respectively. Note that, when the pCP objective is used as dispersion measure the objective (1) is replaced by (8). The resulting model is called CTDP (Center-Based TDP) and it was introduced by Ríos-Mercado and Fernández (2009).

$$z = \max_{i,j \in V} \{d_{ij} x_{ij}\} \tag{8}$$

The NP-completeness of both MTDP and CTDP is well established (Segura-Ramiro et al. 2007; Ríos-Mercado and Fernández 2009). NP-completeness proof for similar models can be found in Altman (1997, 1998) in the context of political districting. For instance, he proved, among others, that the problem of creating equal population/size territories and the problem of redistricting are both NP-complete.

Ríos-Mercado and Fernández (2009) proposed a reactive GRASP for CTDP. Segura-Ramiro et al. (2007) proposed a location-allocation heuristic for MTDP. However, to the best of our knowledge, no exact methods have been developed so far. Although, in theory, the connectivity constraints could be written out explicitly, this would not make any practical sense due to their exponential number. In this work an exact solution procedure to solve both MTDP and CTDP is proposed. In the modeling stage these constraints are not explicitly written and these are generated in an iterative manner within the proposed algorithm. Therefore, the procedure is easily implemented under any algebraic modeler system and it can be solved by any off-the-shelf MILP solver.

Let R_MTDP denote the relaxed model obtained by relaxing (6) from MTDP. In a similar way the relaxed model R_CTDP is defined as the resulting model obtained by relaxing (6) in CTDP. A summary of different MILP relaxed models is displayed in Table 1.

2.2 Integer Quadratic Programming Models

The Integer Quadratic Programming (IQP) model introduced in this work reduces the number of binary variables from n^2 to $2np$. This model is obtained by applying the same technique already used in Domínguez and Muñoz (2008) for a pMP problem and this is the first quadratic formulation for the commercial TDP addressed in this paper. In order to describe the model, a set $Q = \{1, 2, \ldots, p\}$ of territory indices is introduced and binary decision

Table 1 Summary of relaxed models used for solving the MILP and IQP formulations, respectively

Model	Objective	Constraints
R_MTDP	(1)	(2)–(5)
R1_MTDP	(1)	(2)–(5) and (21)
R_CTDP	(8)	(2)–(5)
R1_CTDP	(8)	(2)–(5) and (21)
R_QMTDP	(10)	(11)–(13)
R1_QMTDP	(10)	(11)–(13) and (20)
R_QCTDP	(19)	(11)–(13)
R1_QCTDP	(19)	(11)–(13) and (20)

variables y_{iq} to indicate the territory centers and z_{jq} to represent the assigning of BUs to territories are defined. The parameters are the same as those used in the linear model.

Decision variables for the IQP model

$$z_{jq} = \begin{cases} 1 & \text{if unit } j \text{ is assigned to territory } q;\ j \in V, q \in Q \\ 0 & \text{otherwise} \end{cases}$$

$$y_{iq} = \begin{cases} 1 & \text{if unit } i \text{ is the center of territory } q;\ i \in V, q \in Q \\ 0 & \text{otherwise} \end{cases}$$

According to this definition, the equivalence between the variables in the linear model and the variables in the quadratic model is given by

$$x_{ij} = \sum_{q \in Q} z_{jq} y_{iq}. \tag{9}$$

The resulting IQP model is the following.

$$\text{(QMTDP) minimize } z = \sum_{q \in Q} \sum_{j \in V} \sum_{i \in V} d_{ij} z_{jq} y_{iq} \tag{10}$$

subject to:

$$\sum_{i \in V} y_{iq} = 1 \qquad q \in Q \tag{11}$$

$$\sum_{q \in Q} z_{jq} = 1 \qquad j \in V \tag{12}$$

$$z_{jq} \geq y_{jq} \qquad q \in Q,\ j \in V \tag{13}$$

$$\sum_{j \in V} w_j^{(a)} z_{jq} \geq (1 - \tau^{(a)}) \mu^{(a)} \qquad q \in Q, a \in A \tag{14}$$

$$\sum_{j \in V} w_j^{(a)} z_{jq} \leq (1 + \tau^{(a)}) \mu^{(a)} \qquad q \in Q, a \in A \tag{15}$$

$$\sum_{q \in Q} \sum_{j \in \cup_{v \in S} (N^v \setminus S)} z_{jq} y_{iq} - \sum_{q \in Q} \sum_{j \in S} z_{jq} y_{iq} \geq 1 - |S| \qquad i \in V,\ S \subset [V \setminus (N^i \cup \{i\})] \tag{16}$$

$$z_{jq} \in \{0, 1\} \qquad q \in Q,\ j \in V \tag{17}$$

$$y_{iq} \in \{0, 1\} \qquad q \in Q, i \in V \tag{18}$$

The QMTDP (quadratic median-based territory design problem) model uses an equivalent dispersion measure as that of MTDP. Constraints (11) are to guarantee the location of only one center for each territory. Constraints (12) are for exclusive node assignment. The set of constraints (14) and (15) assure territory balance. Constraints (13) establish that BU j can not be the center of q if j is not assigned to q. According to Proposition 2 in Domínguez and Muñoz (2008), constraints (11) and (12) guarantee the assignment, and constraints (13) are not needed. However, these are shown here for model completeness. The last set of quadratic constraints (16) guarantees connectivity. Again there is an exponential number of these constraints.

Under this quadratic formulation, a dispersion measure based on the pCP objective is given by (19). Let QCTDP (quadratic center-based territory design problem) be the resulting model when the objective function (10) is replaced by the dispersion measure given by (19).

$$\min \quad z = \max_{i,j \in V} \left\{ d_{ij} \sum_{q \in Q} z_{jq} y_{iq} \right\} \tag{19}$$

Note that these IQP formulations are new in the literature for commercial territory design. QMTDP is hard to solve due to the quadratic objective and quadratic connectivity constraints. Additionally, it is not possible to write these explicitly due to its exponential number. If the connectivity constraints are relaxed, the model may be solved using any MINLP method. Let R_QMTDP be the relaxation of QMTDP with respect to the connectivity constraints (16). Clearly, a solution to R_QMTDP provides a lower bound to QMTDP.

There are some special cases for which the model can be strengthened, for example, when there are not feasible solutions containing territories of size 1. In other words, when each feasible solution has territories having at least two basic units associated to it then the following is a valid inequality for R_QMTDP:

$$\sum_{i \in N^j} z_{iq} \geq z_{jq} \quad q \in Q;\ j \in V \tag{20}$$

This condition is true if and only if $w_j^a < (1-\tau^a)\mu^a$ for each $j \in V$, $a \in A$. For our particular case, our data instances always satisfy this condition, and therefore the model can make use of these valid inequalities. These inequalities can be interpreted as follows. If j is assigned to territory q at least one of its neighbors ($i \in N^j$) must be assigned to the same territory. In this sense, these constraints avoid the unconnected subsets S with $|S| = 1$. The motivation for this stems from the fact that empirical work showed that a very large proportion of (unconnected) optimal solutions to the relaxed models R_MTDP, R_CTDP, R_QMTDP, or R_QCTDP come from subsets of cardinality equal to 1. Given that there is a polynomial number of these kind of subsets their related connectivity constraints can be easily incorporated into the model.

Note that, for MILP formulations the equivalent valid inequalities are given by:

$$\sum_{l \in N^j} x_{il} \geq x_{ij} \quad i \in V, \quad j \in V \setminus (\{i\} \cup N^i) \tag{21}$$

In contrast with valid inequalities (20) that are valid only when the condition of no singleton territories hold, constraints (21) are valid for any instance.

Let R1_QMTDP be the relaxation defined by R_QMTDP plus the additional constraints (20). In a similar way we can define the relaxed models for the QCTDP model. These are called R_QCTDP and R1_QCTDP, respectively. Similarly, for both MTDP and CTDP models, new relaxed models are obtained by adding (21) in the relaxed models R_MTDP and R_CTDP, respectively. We called these R1_MTDP and R1_CTDP, respectively. For better definition of these relaxed models see Table 1.

In the following section we outline a solution framework that can be used to solve any of these models. This procedure can be used to solve the problem using both MILP and IQP formulations. This procedure guarantees a global optimal solution for MILP models and local or global optimal solutions for IQPs, depending on what method is used for solving the relaxed subproblem.

3 Solution procedure

One of the main difficulties for obtaining exact solutions for any of these models arise from the exponential number of connectivity constraints. The explicit enumeration of these constraints results practically impossible. Thus, to get optimal solutions we devise an iterative procedure that uses branch and bound (B&B) and a cut generation scheme.

The idea is relatively simple. By relaxing the connectivity constraints, we are left with a relaxed problem that is solved by B&B. Then, the solution to this relaxed problem is checked for connectivity. The connectivity test is done by solving a separation problem (Algorithm 2) that is polynomially solvable throught the breadth first search (BFS) algorithm. The corresponding identified violated valid inequalities (if any) are then added to the relaxed model as cuts and the procedure continues until no more violated inequalities are found. The Iterative Cut Generation Procedure for solving TDPs (ICGP-TDP) is outlined in Algorithm 1. For solving the MILP relaxed models, the SolveMILP method in ICGP-TDP uses any branch-and-bound method. In contrast, the SolveIQP method may call either an exact or an approximate method. In our case, we are attempting to come up with a way to find faster solutions, so we make use of a local optimum method for finding attractive feasible solutions for the IQP relaxed models. An issue to be investigated is precisely the trade-off between time and solution quality. Assuming a finite algorithm is used for solving the integer relaxed models (in SolveMILP() or

Algorithm 1 ICGP-TDP(P, $DispMeasure$, $ModelType$)

Input:
P:=Instance of the TDP problem
$DispMeasure$:= pCP or pMP objective function
$ModelType$:= MILP or IQP
Output: $X = (X_1, X_2, \ldots, X_p)$:= A feasible p-partition of V
$Cuts \leftarrow \emptyset$ {*Cut set*}
$Model \leftarrow$ GenerateRelaxedModel(P, $DispMeasure$, $ModelType$)
While($Cuts \neq \emptyset$)
  If($ModelType$ = MILP)
  $X \leftarrow$ SolveMILP($Model$)
  Else
  $X \leftarrow$ SolveIQP($Model$)
  EndIf
  $Cuts \leftarrow$ SolveSeparationProblem(P,X)
  AddCuts($Model$, $Cuts$)
EndWhile
Return X

SolveIQP()), the convergence of the algorithm is guaranteed due to the fact that the separation problem is solved exactly (in polynomial time) returning either a set of violated connectivity constraints or an empty set. Because there is a finite set of connectivity constraints the algorithm is guaranteed to stop. When it stops, the last solution is feasible with respect to the connectivity constraints, and therefore, an optimal solution to the problem.

3.1 The separation problem

Suppose we have a graph $G(V, E)$ and a p-partition $X = (X_1, X_2, \ldots, X_p)$, where each of these sets $X_k, k = 1, \ldots p$ induces a subgraph $G_k = (X_k, E(X_k))$ of G and a center $c_k \in X_k$. The separation problem consists of first identifying all connected components of G_k. This can be done very efficiently by breadth-first-search (BFS) in $O(|E|)$ as follows. Starting from any node i BFS (Cormen et al. 1990) is applied to find a node j adjacent or reachable from i. This is repeated until no more nodes can be reached from the previously formed node set. At this point, this node set is one connected component of G and we proceed iteratively starting from any other non-visited node. The procedure stops when all nodes have been visited. BFS assures that this is accomplished in $O(|E|)$. Then, for each $k = 1, \ldots, p$, each of the connected components

Algorithm 2 SolveSeparationProblem(P, X)

Input:
P:= Instance of the TDP problem
$X = (X_1, X_2, \ldots, X_p)$:= A p-partition of V
Output: *Cuts* := Set of violated connectivity constraints
Cuts $\leftarrow \emptyset$
For $(k = 1, \ldots, p)$

Obtain connected components $S_1, S_2, \ldots, S_t$ of $G(X_k, E(X_k))$
For each S_t such that $c_k \notin S_t$ generate the violated cut and add it to *Cuts*

EndFor
Return (*Cuts*)

of G_k that does not contains the center c_k is used to generate a violated connectivity constraint in our problem. In other words, each component S_t of G_k plays the role of set S in constraints (6). Algorithm 2 describes the steps to solve the separation problem. Note that in our implementation, the BFS algorithm is used for obtaining the connected components. This algorithm runs in polynomial time.

In order to illustrate the IGCP-TDP algorithm consider an example with $n = 11$ nodes and $p = 2$ territories (see the graph in Fig. 1). Suppose that the solution to the relaxed problem (without connectivity constraints) after applying branch and bound is depicted in Fig. 1,where the dotted lines collect BUs belonging to the same territory and the nodes 4 and 5 are the territory centers. This solution corresponds to the 2-partition given by $X_1 = \{1, 4, 6, 7, 11\}$

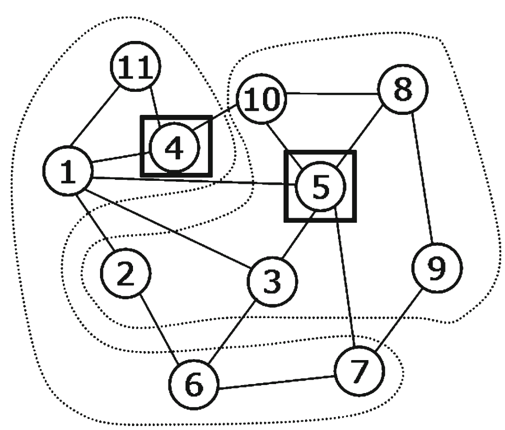

Fig. 1 Example of an unconnected territory design for p=2 and n=11

with center in $c_1 = 4$ and $X_2 = \{2, 3, 5, 8, 9, 10\}$ with center in $c_2 = 5$, then the variables have the following values:

$$x_{41} = x_{44} = x_{46} = x_{47} = x_{411} = 1,$$
$$x_{42} = x_{43} = x_{45} = x_{48} = x_{49} = x_{410} = 0,$$
$$x_{52} = x_{53} = x_{55} = x_{58} = x_{59} = x_{510} = 1,$$
$$x_{51} = x_{54} = x_{56} = x_{57} = x_{511} = 0;$$
$$x_{ij} = 0, \forall i, j \in V, i \neq \{4, 5\}.$$

Given this solution, the separation problem (Algorithm 2) is solved to identify those connectivity constraints (6) that are violated by this solution. Applying the BFS algorithm, the connected components $S_1, \ldots, S_t$ are identified on each territory. As can be seen from X_1 the connected component $S_1 = \{6, 7\}$ is unconnected from c_1 then it induces a violated constraint which is generated as

$$x_{42} + x_{43} + x_{45} + x_{49} - x_{46} - x_{47} \geq -1$$

Similarity, in X_2 the connected component $S_1 = \{2\}$ is unconnected from c_2 and the violated constraint is given by

$$x_{51} + x_{56} - x_{52} \geq 0$$

Following the ICGP-TDP procedure, we add the cuts (violated constraints) to the relaxed model and it is solved again. We proceed iteratively until the final solution gives us a connected territory design or a feasible solution is not found. The latter means that the original problem is infeasible.

Note that, in each iteration, the number of aggregated cuts is equal to the total number of unconnected subsets identified from the given solution. Within a given iteration this number of identified unconnected subsets is bounded by n; however, in the worst case the total number of these identified subsets during the execution of the algorithm could be exponential. Nonetheless, in practice this number is found to be relatively low as it will be seen in the following section.

4 Computational results

The proposed ICGP-TDP method was coded in C++ and compiled with the Sun C++ 8.0 compiler. The MILP relaxations are solved through CPLEX 11.2 and the IQP relaxations are solved by DICOPT, one of the most popular methods for solving non-linear mixed-integer programs developed by J. Viswanathan and Ignacio E. Grossmann at the Engineering Design Research Center (EDRC) at Carnegie Mellon University (see Kocis and Grossmann (1989) and Viswanathan and Grossmann (1990) for more details). Two stopping criteria were used: by optimality gap ($gap \leq 5 \times 10^{-6}$) and by time (7,200 s). In order to speed up convergence, priorities were used on

Table 2 Results of ICGP-TDP when applied to CTDP under the R_CTDP relaxation

Size	Solved at	Iterations		Solved	Cuts/inst		Time (s)	
(n, p)	1st iter (%)	Ave	Max	(%)	Ave	Max	Ave	Max
(60, 4)	20	5.3	26	100	12.1	82	381	1,446
(80, 5)	10	5.4	14	90	12.4	43	2,682	7,200
(100, 6)	10	2.3	11	40	3.5	32	5,812	7,200
(150, 8)	0	–	–	0	–	–	7,200	7,200

the binary variables to ensure that x_{ii} are branched before than $x_{ij}, i \neq j$, $i, j \in V$. Randomly generated instances based on real-world data provided by the industrial partner were used. Each instance topology was generated by using the generator developed by Ríos-Mercado and Fernández (2009). In this work, the authors used historical information from the firm and obtained the data distribution associated to the number of customers and sales volume. The firm uses Euclidean distances between basic units as computed from their GIS. We considered a tolerance $\tau^{(a)} = 0.05, a \in A$, and generated three different instance sets as $(n, p) \in \{(60, 4), (80, 5), (100, 6)\}$. For each of these sets, 20 different instances were generated. Additionally, ten different instances of two larger sets were generated for $(n, p) \in \{(150, 8), (200, 11)\}$. The codes and data sets are available at http://yalma.fime.uanl.mx/~roger/ftp/tdp/.

4.1 Evaluation of MILP models

We first evaluate linear models CTDP and MTDP when the relaxed models R_CTDP and R_MTDP, respectively, are used within the ICGP-TDP procedure.

Tables 2 and 3 show the results for CTDP and MTDP, respectively. The first column indicates the instance size tested. The second column shows the percentage of instances that were solved at the first iteration (out of 20 except for the set (150, 8)), that is, the percentage of instances for which a connected partition was found at the first iteration. The third column contains the average and the maximum number of iterations per instance required by the algorithm to find the optimal solution. The fourth column displays the percentage of instances solved within the specified time limit. The fifth column shows the average and the maximum number of cuts added per instance solved.

Table 3 Results of ICGP-TDP when applied to MTDP under the R_MTDP relaxation

Size	Solved at	Iterations		Solved	Cuts/inst		Time (s)	
(n, p)	1st iter (%)	Ave	Max	(%)	Ave	Max	Ave	Max
(60, 4)	80	1.4	6	100	0.5	5	7	33
(80, 5)	70	1.4	4	100	0.5	4	53	235
(100, 6)	75	1.4	4	100	0.5	4	95	438
(150, 8)	75	1.8	5	80	1.6	6	1,900	7,200

Table 4 Size of unconnected subsets for the R_CTDP relaxation

Size	Cuts	% Cuts with			
(n, p)	identified	$\lvert S\rvert = 1$	$\lvert S\rvert = 2$	$\lvert S\rvert = 3$	$\lvert S\rvert \geq 4$
(60, 4)	44	72.7	18.3	4.5	4.5
(80, 5)	65	58.5	20	7.7	13.8
(100, 6)	103	67	11.6	7.8	13.6
(150, 8)	–	–	–	–	–

Finally, the last column displays information about the CPU time (average and maximum) used per instance.

For model CTDP, Table 2 indicates that a very small proportion of the instances were solved at the first iteration. As many as 26 iterations and 82 cuts were needed in the worst case to solve the problems of size (60, 4). At the end of the procedure, all instances of the (60, 4) were solved optimally. 90% of the (80, 5) set were solved optimally. However, the procedure struggled with the larger sets. For the two smaller sets, around five iterations and 12 cuts were needed on average. Note that, for a specific iteration the separation problem has the property to identify more than one unconnected subset and it generates all violated connectivity constraints at the same iteration. Note that for the (150, 8) set, the procedure was unable to terminate a single iteration within the time limit.

These statistics improve significantly for the MTDP model (Table 2). Except for a very few cases in the largest set, all other instances were solved optimally. A large proportion of these were solved at the very first iteration. On average, this required less than 2 iterations and a very few cuts for obtaining optimal solutions. This suggests not only that the LP relaxation of the median-based model is tighter that the one of the center-based model, but also that solutions to the R_MTDP relaxation yield near-connected solutions. This has a positive impact on the overall solution time.

Another issue to investigate is to whether or not the introduction of constraints (21) has a positive effect on strengthening the model. Recall that constraints (21) eliminate unconnected subsets of size 1. Thus, in this experiment we solved the very first relaxation only for every instance and tallied the cardinality of all unconnected subsets for both CTDP and MTDP. A summary of this experiment is shown in Tables 4 and 5. As we can see in Table 4, most of the identified cuts for CTDP correspond to unconnected subsets of cardinality equal to 1. For the (60, 4), (80, 5), and (100, 6) sets, the

Table 5 Size of unconnected subsets for the R_MTDP relaxation

Size	Cuts	% Cuts with		
(n, p)	identified	$\lvert S\rvert = 1$	$\lvert S\rvert = 2$	$\lvert S\rvert \geq 3$
(60, 4)	4	100	0	0
(80, 5)	6	83	17	0
(100, 6)	5	80	20	0
(150, 8)	6	83	17	0

Table 6 Comparison of relaxations R_MTDP and R1_MTDP

Size (n, p)	Solved at 1st iteration (%)		Cuts added		Solved (%)	
	R	R1	R	R1	R	R1
(60, 4)	80	100	9	0	20	100
(80, 5)	70	95	9	1	20	100
(100, 6)	75	90	9	2	20	100
(150, 8)	5	45	6	3	80	90

proportion of unconnected subsets of cardinality 1 is 72.7, 58.5, and 67.0%, respectively. This proportion is even more dramatic for MTDP (see Table 5). One can see that the number of total unconnected subsets is considerable smaller than that of the R_CTDP relaxation. This confirms that the MTDP model not only has a better LP relaxation, but it also favors connectivity, which is a very important issue. Hence, these results clearly justify and motivates the introduction of the valid inequalities given by (21) into the relaxed models.

The following experiment clearly illustrates this issue. We now solve model MTDP under two different relaxations: R_MTDP and R1_MTDP (incorporating the valid inequalities). We identify these as R and R1, respectively. Table 6 displays the results. The second and third columns show the number of instances (out of 20) that were solved optimally at the very first iteration, that is, by solving the first relaxed models for R_MTDP and R1_MTDP, respectively. The fourth and fifth columns display the total number of cuts added during the execution of the algorithm. The last two columns show the percentage of instances that were optimally solved. As can bee seen, relaxation R1 provides a more attractive choice in all senses. Therefore, the introduction of constraints (21) into the relaxed model provides a stronger LP representation of model MTDP. This has indeed a positive impact in solution times.

4.2 Evaluation of IQP models

We now consider the IQP formulations QCTDP and QMTDP under the R_QCTDP and R_QMTDP relaxations, respectively. In a similar fashion as carried out with the linear models, we investigate the distribution of the cardinality of the unconnected subsets when only the very first relaxation is solved. Tables 7 and 8 display the results for QCTDP and QMTDP, respectively.

Table 7 Size of unconnected subsets for the R_QCTDP relaxation

Size	Cuts	% Cuts with			
(n, p)	identified	$\|S\| = 1$	$\|S\| = 2$	$\|S\| = 3$	$\|S\| \geq 4$
(60, 4)	662	68	21	6	5
(80, 5)	956	73	17	6	4
(100, 6)	1,340	77	17	4	2
(150, 8)	1,088	82	14	3	1

Table 8 Size of unconnected subsets for the R_QMTDP relaxation

Size	Cuts	% Cuts with		
(n, p)	identified	$\|S\| = 1$	$\|S\| = 2$	$\|S\| \geq 3$
(60, 4)	3	100	0	0
(80, 5)	5	40	20	40
(100, 6)	6	67	33	0
(150, 8)	5	60	40	0

The description is similar to that of Table 4. It can be seen that most of the unconnected subsets have cardinality 1, which is a behavior also observed in the linear models. Another observation is that the relaxation of the median-based model provides solutions with a higher degree of connectivity that the one provided by the center-based model. Hence a considerable less amount of effort will be needed to eventually solved a median-based model with connectivity constraints. These results clearly motivate the introduction of valid inequalities (20) into the relaxed models.

We now evaluate the effect of incorporating constraints (20) into the relaxed R1_QMTDP model. Table 9 shows the results when QMTDP is solved under the corresponding R1_QMTDP relaxation. The second column shows the percentage of instances that were solved at the very first iteration. The third column displays the total average number of cuts added. Columns 4 through 6 gives information on the number of iterations needed to reach optimality. As it can be seen, the addition of constraints (20) gives very competitive results as little additional effort was needed for cut generation, with a small number of iterations.

When attempting to carry out a similar experiment for the QCTDP model under the corresponding R1 relaxation, it was observed that the LP relaxation was still extremely weak. The procedure could not terminate a single iteration within the specified time limit. The effect of adding the cuts resulted in even higher running times. Thus, clearly effort did not result in a satisfactory payoff.

4.3 Comparing MILP and IQP

Clearly, we have seen that solving the quadratic models is faster than solving the linear models. However, solving the quadratic model with local-optimum methods no longer assures global optimality. Therefore, an important issue to be investigated is precisely the trade-off between solution quality and

Table 9 Solution of QMTDP under the R1_QMTDP relaxation

Size	Solved at	Cuts	Iterations		
(n, p)	1st iter (%)	added	Min	Ave	Max
(60, 4)	95	1	1	1.1	2
(80, 5)	85	4	1	1.2	2
(100, 6)	95	1	1	1.1	2
(150, 8)	100	0	1	1.0	1

Table 10 Comparison of MTDP and QMTDP models for instances in the set (60, 4)

Inst	Objective value		Gap (%)	Time (s)	
	MTDP	QMTDP		MTDP	QMTDP
1	5,305.57	5,306.00	0.01	4	2
2	5,451.68	5,463.00	0.21	4	2
3	5,507.88	5,553.00	0.82	10	2
4	5,935.67	6,114.00	3.00	4	6
5	5,303.20	5,303.20	0.00	3	2
6	5,253.94	5,280.00	0.50	33	3
7	5,460.18	5,855.00	7.23	4	3
8	5,309.96	5,314.00	0.08	4	2
9	5,224.51	5,225.00	0.01	2	3
10	5,350.15	6,140.00	14.76	3	2
11	5,150.91	5,152.00	0.02	3	2
12	5,597.50	5,705.00	1.92	6	2
13	5,731.99	5,732.00	0.00	3	3
14	5,462.96	5,869.00	7.43	5	2
15	5,332.77	5,759.00	7.99	6	2
16	5,399.54	5,499.00	1.84	14	2
17	5,602.86	5,603.00	0.00	3	2
18	5,773.96	6,299.00	9.09	4	4
19	5,543.45	5,544.00	0.01	17	2
20	5,767.54	5,768.00	0.01	4	2

computational effort. We apply the solution procedure to models MTDP and QMTDP on 20 instances of data sets {(60, 4), (80, 5), (100, 6)} and ten instances of data set (150, 8). Detailed results for instances of (60, 4) and (150, 8) are shown in Tables 10 and 11, respectively. The fourth column shows the relative optimality gap of the solution found under the quadratic model (that is, with respect to the optimal solution found by the linear model). For the instances marked with a star, the MILP could not find an optimal solution within the specified time limit so a best integer solution is used instead.

As can be seen from Table 10, for 19 out of 20 instances the solution found with the quadratic model falls within 10% of the optimal solution, and 60% of the solutions lay within 1% of optimality. Time is not an issue in these sets as can be seen in the last two columns. However, for the larger instances (displayed in Table 11), time becomes important. We can see how

Table 11 Comparison of MTDP and QMTDP models for instances in the set (150, 8)

Inst	Objective value		Gap (%)	Time (s)	
	MTDP	QMTDP		MTDP	QMTDP
1	9,511.76	9,979	4.91	1,137	9
2	9,404.60 (*)	9,509	1.11	7,200	29
3	9,125.61	9,130	0.05	90	32
4	9,359.00	9,646	3.07	147	30
5	9,506.58	10,494	10.39	455	42
6	9,039.06	9,088	0.54	78	25
7	9,819.18	10,017	2.02	1,842	29
8	9,202.13 (*)	9,550	3.78	7,200	34
9	9,670.90	9,972	3.11	730	28
10	9,570.58	9,794	2.33	125	26

Table 12 Time comparison for QMTDP and MTDP models

Size	MTDP time (s)			QMTDP time (s)		
(n, p)	Min	Ave	Max	Min	Ave	Max
(60, 4)	2	6.8	33	2	2.5	6
(80, 5)	8	53.2	235	4	5.6	12
(100, 6)	18	94.8	438	7	8.8	23
(150, 8)	78	1,900.4	7,200	9	28.4	42

time significantly increases for the MILP model. There are two instances where time limit was reached when using the MILP model. When using the quadratic model, all instances were solved within 1 minute of CPU time, delivering optimality gaps of less than 5% in 90% of the instances. Thus, this makes the quadratic model a very attractive choice for relatively large instances.

A summary of the comparison between MTDP and QMTDP models are displayed in Tables 12 and 13. The computational effort is shown in Table 12 and the solution quality over four different data sets is shown in Table 13. As we can see from these tables, CPU time employed for solving the quadratic model is relatively low compared with the time used by the linear model. Furthermore, the average relative optimality gaps for the quadratic model are less than 4%. In many cases the solution to the quadratic model was less than 1%.

In addition, we attempted to solve ten instances of size (200, 11) by using both MTDP and QMTDP models. The ICGP-TDP procedure was able to produce the optimal solution for four instances (using the MTDP model). In contrast, by using the QMTDP model, the ICGP-TDP procedure reported locally optimal solutions for nine out of ten instances within the specified time limit (7,200 s). Table 14 displays each one of these large instances, where mark (*) is used to identify those cases in which the optimization stopped by time limit. For those cases, the best integer solution found is used to make the comparison. Observe that the percentage of relative optimality between the MTDP and QMTDP solutions is in the worst case equal to 10.56% and in the best case it is equal to 1.15%. That means, the IQP formulation proposed in this paper allows to solve larger instances than the MILP formulation by using shorter optimization time.

Additionally, an instance with $\tau^{(a)} = 0.05, a \in A$; $n = 280$ and $p = 9$ was generated. This instance was tested using the MTDP formulation and in the first relaxed model (R_MTDP), the branch and bound reported a percentage of relative optimality equal to 14.68%, after 2 hours. This percentage of gap is computed by $[(BestInt(R\_MTDP) - BestLB(R\_MTDP))/BestInt$

Table 13 Solution quality for QMTDP

Size	Gap (%)		
(n, p)	Min	Average	Max
(60, 4)	0.00	2.75	14.80
(80, 5)	0.01	2.61	8.15
(100, 6)	0.06	3.14	7.56
(150, 8)	0.05	3.13	10.39

Table 14 Comparison of MTDP and QMTDP models for instance set (200, 11)

Inst	Objective value		Gap (%)	Time (s)	
	MTDP	QMTDP		MTDP	QMTDP
1	10,422.01	11,523	10.56	1,116	28
2	10,646.14 (*)	11,425	7.32	7,200	966
3	10,846.77	11,443	5.50	1,468	7,200
4	11,122.03 (*)	11,443	2.89	7,200	3,618
5	10,878.12 (*)	11,097	2.01	7,200	1,193
6	10,499.29 (*)	10,746	2.35	7,200	1,871
7	11,061.00 (*)	11,686	5.65	7,200	1,088
8	10,659.51	11,205	5.12	2,641	592
9	11,470.29 (*)	11,648	1.55	7,200	1,263
10	11,043.82	11,780	6.67	1,211	2,349

$(R\_MTDP)] \times 100)$. The same instance was tested by using R1_QMTDP and ICGP-TDP reported a connected solution in less than 4 min. Comparing the objective value for QMTDP with the best lower bound found by branch and bound, we computed a relative optimality of 7.15%. The percentage of gap (relative optimality) was computed by $[(Best(QMTDP) - BestLB(MTDP))/Best(QMTDP)] \times 100$.

Finally, even the ICGP-TDP procedure (with QMTDP model) was tested for an instance with $n = 500$ and $p = 12$ resulting in a locally optimal solution for QMTDP without reaching the time limit of 2 h (7,200 s). The objective value associated to this connected solution is equal to 27,113.42. In contrast, using the MTDP formulation within the ICGP-TDP procedure we observed that it stopped by time during the optimization of the first relaxed model (R_MTDP). The best integer solution reported by B&B has an objective value equal to 38,905.19 with a gap equal to 42.90%. Note that this solution is the best found solution for the problem without connectivity constraints.

We conclude that the QMTDP model is a fast and attractive alternative to find relatively good solutions also for large instances because it offers a good compromise between time and quality.

5 Conclusions

In this work we have proposed new IQP models for the commercial territory design problem with connectivity and multiple balancing constraints. These IQP formulations use a significantly smaller number of binary variables. In addition, we have developed an exact solution procedure (ICGP-TDP) based on branch and bound and a cut generation strategy. The method can be applied to both MILP and IQP models. This is the first exact algorithm developed to date for this problem. The models were strengthen by the introduction of valid inequalities that eliminate unconnected subsets of size 1. We have observed empirically that most of the unconnected subsets found in the relaxed models (relaxing the connectivity constraints) have cardinality equal to 1, so this motivates the introduction of these valid inequalities. We empirically proved that the cut did in fact helped to find connected territories faster.

When the solution method was applied to solve instances under the linear and quadratic models, the proposed IQP models showed a balanced performance between quality and effort. For the larger instances, execution times under the quadratic models were significantly lower that those observed under the linear models. The solution quality of those obtained by the quadratic model over all instances was in the range of 0.0–14.8%, and in most cases, less than 5%.

We observed that the *p*MP objective is more LP-friendly than the *p*CP objective. During the branch and bound process the linear relaxation for *p*MP objective showed better performance than the linear relaxation for the *p*CP objective. Furthermore, it was also observed that solutions obtained from the relaxation of the median-based models had a very high degree of connectivity. This had a very good impact on computational efficiency since very few iterations were needed to find connected solutions as opposed to the center-based models. Therefore, in the absence of a standard dispersion measure, the *p*MP objective may be a good choice for other territory design problems that have compactness as performance measure.

In this work, we efficiently solved instances with up to 150 BUs and eight territories using MILP models. Literature review in territory design shows that the largest instance with connectivity constraints solved optimally had no more than 50 BUs (Garfinkel and Nemhauser 1970). As far as this particular commercial TDP is concerned, our proposed method is the first exact optimization scheme developed for the problem. For IQPs models, we obtained locally optimal solutions for instances with up to 500 BUs and 12 territories. This instance size is intractable under MILP formulations. One of the advantages of the proposed approach is that it can be implemented relatively easy with off-the-shelf MILP and IQP solvers.

There are several extensions to this problem that deserve attention. For instance, this work is based on using Euclidean distances to represent distances between cities. While it is true that in location problems replacing Euclidean by network or shortest path distances can be done without loss of generality, this cannot be done in this type of territory design problems due to the presence of the connectivity constraints. It turns out that shortest path distances between units are solution dependent because this shortest path must belong entirely to the same territory. This makes the problem a lot more difficult to solve. As seen in literature, the compactness measures based on Euclidean distances provide a relatively good choice in the design stage. It is clear, however, that using network distances would become more relevant in a posterior routing stage.

Acknowledgements This work was supported by the Mexican National Council for Science and Technology, grant SEP-CONACYT 48499-Y, and Universidad Autónoma de Nuevo León under its Scientific and Technological Research Support Program, grants UANL-PAICYT CA1478-07 and CE012-09. The research of the first author was funded by a doctoral fellowship from Universidad Autónoma de Nuevo León, grant NL-2006-C09-32652. The authors would like to thank three anonymous referees for their comments and suggestions which have significantly improved this paper.

References

Aerts JCJH, Eisinger E, Heuvelink GBM, Stewart TJ (2003) Using linear integer programming for multi-site land-use allocation. Geogr Anal 35(2):148–169

Altman M (1997) The computational complexity of automated redistricting: is automation the answer? Rutgers Comput Technol Law J 23(1):81–142

Altman M (1998) Districting principles and democratic representation. PhD thesis, California Institute of Technology, Pasadena

Bozkaya B, Erkut E, Laporte G (2003) A tabu search heuristic and adaptive memory procedure for political districting. Eur J Oper Res 144(1):12–26

Caballero-Hernández SI, Ríos-Mercado RZ, López F, Schaeffer E (2007) Empirical evaluation of a metaheuristic for commercial territory design with joint assignment constraints. In: Fernandez JE, Noriega S, Mital A, Butt SE, Fredericks TK (eds) Proceedings of the 12th annual international conference on industrial engineering theory, applications, and practice (IJIE). Cancun, Mexico, pp 422–427. ISBN: 978-0-9654506-3-8

Cormen TH, Leiserson CE, Rivest RL (1990) Introduction to algorithms. MIT Press, Cambridge

Domínguez E, Muñoz J (2008) A neural model for the *p*-median problem. Comput Oper Res 35(2):404–416

Drexl A, Haase K (1999) Fast approximation methods for sales force deployment. Manage Sci 45(10):1307–1323

Fleischmann B, Paraschis JN (1988) Solving a large scale districting problem: a case report. Comput Oper Res 15(6):521–533

Garfinkel RS, Nemhauser GL (1970) Solving optimal political districting by implicit enumeration techniques. Manage Sci 16(8):B495–B508

Hess SW, Samuels SA (1971) Experiences with a sales districting model: criteria and implementation. Manage Sci 18(4):998–1006

Hess SW, Weaver JB, Siegfeldt HJ, Whelan JN, Zitlau PA (1965) Nonpartisan political redistring by computer. Oper Res 13(6):998–1006

Hojati M (1996) Optimal political districting. Comput Oper Res 23(12):1147–1161

Kalcsics J, Nickel S, Schröder M (2005) Towards a unified territorial design approach: applications, algorithms, and GIS integration. Top 13(1):1–56

Kocis GR, Grossmann IE (1989) Computational experience with DICOPT solving MINLP problems in process systems engineering. Comput Chem Eng 13(3):307–315

Marlin P (1981) Application of the transportation model to a large-scale districting problem. Comput Oper Res 8(2):83–96

Mehrotra A, Johnson EL, Nemhauser GL (1998) An optimization based heuristic for political districting. Manage Sci 44(8):1100–1113

Ricca F, Simeone B (2008) Local search algorithms for political districting. Eur J Oper Res 189(3):1409–1426

Ríos-Mercado RZ, Fernández EA (2009) A reactive GRASP for a commercial territory design problem with multiple balancing requirements. Comput Oper Res 36(3):755–776

Segura-Ramiro JA, Ríos-Mercado RZ, Álvarez-Socarrás AM, de Alba Romenus K (2007) A location-allocation heuristic for a territory design problem in a beverage distribution firm. In: Fernandez JE, Noriega S, Mital A, Butt SE, Fredericks TK (eds) Proceedings of the 12th annual international conference on industrial engineering theory, applications, and practice (IJIE). Cancun, Mexico, pp 428–434. ISBN: 978-0-9654506-3-8

Shirabe T (2009) Districting modeling with exact contiguity constraints. Environ Plann B Plann Des 36(6):1053–1066

Viswanathan J, Grossmann IE (1990) A combined penalty function and outer approximation method for MINLP optimization. Comput Chem Eng 14(7):769–782

Xiao N (2006) An evolutionary algorithm for site search problems. Geogr Anal 38(3):227–247

Zoltners AA, Sinha P (2005) Sales territory design: Thirty years of modeling and implementation. Mark Sci 24(3):313–331