An Extendable Heuristic Framework to Solve the *P*-Compact-Regions Problem for Urban Economic Modeling

Wenwen Li , Richard Church, Michael Goodchild

Abstract

The *p*-compact-regions problem involves generating a fixed number (*p*) of regions from *n* atomic polygonal units with the objective of maximizing the compactness of each region. Compactness is a shape factor measuring how closely and firmly the polygonal units in a region are packed together. A compact polygonal region has the advantages of being homogeneous and maximizing the accessibility of all points within that region, therefore it is useful in a large number of real-world applications, such as in conservation planning, political district partitioning, and the proposed application in this paper concerning regionalization for urban economic modeling. This paper reports our efforts in designing a heuristic framework that involves semi-greedy growth and local search to solve the *p*-compact-regions problem to optimality or near-optimality. We apply this model to support urban economic simulation, in which activities need to be aggregated from the 4,109 Transportation Analysis Zones (TAZs) of six southern California counties into 100 regions to achieve desired computational feasibility of the economic simulation model. Spatial contiguity, physiography, political boundaries, the presence of local centers, and intra-zonal and inter-zonal traffic are considered, and efforts are made to ensure consistency of selected properties between the disaggregated and aggregated regions. This work makes an original contribution in the development of a highly extendable and effective solution framework to allow researchers to investigate large, real, non-linear regionalization problems and find practical solutions.

Keywords: Spatial optimization, greedy, heuristic, compactness, clustering, moment of inertia, simulated annealing, TABU, GRASP, regionalization, zoning, clustering

1. Introduction

The *p*-compact-regions problem involves aggregating *n* spatially contiguous polygonal units into *p* regions ($p \leq n$) while maximizing each region's compactness, potentially optimizing other defined objectives and satisfying the constraint that the contiguity of each resulting polygonal region must be preserved (Li et al. 2014). This problem differs from traditional clustering algorithms, such as centroid-based models (*K*-means) or density models (DBSCAN: Density-Based Spatial Clustering of Applications with Noise), in that the set of data to cluster in a *p*-compact-regions problem is a subdivision of a larger continuous surface, such as clustering all census tract data within the state of Arizona into *p* regions, rather than a dataset that can be modeled as point data, such as the type of dataset used in *K*-means. The *p*-compact-regions problem is indeed an optimization problem, involving one or more predefined objectives, such as minimizing dissimilarity among basic units within a cluster/region and maximizing the compactness of the resulting regions. In clustering polygonal spatial objects, shape and contiguity become essential factors to consider in the algorithmic design.

Compactness of a polygonal object is usually considered an important indicator of shape in regionalization problems. Compactness is acknowledged as one of the most intriguing properties of a shape (Angel et al. 2010) because a compact region is likely to be homogeneous, sharing common attributes and properties (Li et al. 2013a). The contiguity constraint for the *p*-compact-regions problem, on the other hand, ensures that each pair of polygonal units within the same region can be connected by a path which falls entirely within the region. According to the literatures of computer science, geographic information science (GIScience), and regional science, this type of clustering can also be termed zonation, districting, and regionalization—all of which apply to many domains (Duque et al. 2011). The applications include discovering satisfactory coverage of emergency medical service facilities for first aid (Gendreau et al.2005; Alsalloum and Rand 2006; Sorensen and Church 2010); identifying regions that maximize the coverage of selected species in conservation planning (Church et al. 1996); partitioning an area into compact electoral districts to avoid political gerrymandering (Young 1988; Pang et al. 2010); and designing compact sales territories to maximize profit and minimize costs of customer service (Hess and Samuels 1971).

In this paper, we propose a heuristic-based solution framework for a *p*-compact-regions problem designed for determining the planning units in a microeconomic model of urban land use and transportation called RELU-TRAN (Alex and Liu 2007). RELU-TRAN is a dynamic general-equilibrium model of a metropolitan economy and its uses of land. It equilibrates floor space, land and labor markets, and the market for the products of industries, treating development (construction and demolition), spatial inter-industry linkages, commuting, and discretionary travel. Mode choice and equilibrium congestion on the highway network is treated by integrating an algorithm involving stochastic user equilibrium. To facilitate public policy analysis in the greater Los Angeles metropolitan region, researchers from SUNY Buffalo, University of California Santa Barbara, University of California Riverside, and Arizona State University have formed a collaborative team to revolutionize the use of this computable economic model (Li et al. 2013b).

One fundamental task of this research effort is to generate model zones for numerical simulation from 4,109 Traffic Analysis Zones (TAZs) to achieve the computational feasibility of the RELU-TRAN model for the prediction of urban changes in Southern California. Model zone compactness is considered a primary goal because it implies maximum accessibility of all parts within each zone. Meanwhile, factors including spatial contiguity, the coincidence of model-zone boundaries with physiographic features and political boundaries, and the patterns of intra-zonal and inter-zonal traffic need to be considered. This is a "supreme" problem for three reasons. First, it requires iterative computation among more than 4,000 atomic units, which is large in comparison with the literature on similar problems. Second, it needs to address many constraints. Third, the objective function is complex and non-linear.

To solve this particular *p*-compact-regions problem to near-optimality, we propose an efficient heuristic framework that is extendable to integrate a variety of popular heuristic approaches and compactness measures for the purpose of identifying the best configuration. The following sections provide detailed discussions on previous work (Section 2), formulation of the real-world *p*-compact-regions problem in the context of urban economic modeling (Section 3), design and implementation of the heuristic solution framework (Section 4 and 5), and a series of experiments to demonstrate the performance of the proposed method (Section 6). In Section 7, we conclude the work and discuss future research directions.

2. Literature

Over the past several decades, researchers have developed a number of methods for providing solutions to versions of the *p*-regions problem (Cova and Church, 2000). Optimal solutions for versions requiring contiguous but non-compact zones have been obtained using integer programming and integer goal programming. However, when compactness is a requirement the computational complexity and the non-linear nature of the *p*-regions problem makes it difficult to formulate and solve using general-purpose optimization software.

Fischer and Church (2003) used an indirect measure of compactness, minimizing the length of the outside perimeter of a selected set of areas, as a means to cluster selected units and encourage compactness. This was accomplished using an integer-linear optimization model. This method provides only a rough estimation of compactness. A narrow and long rectangle and a square which have the same perimeter would be considered equally compact, but in fact, the square shape is much more compact than the rectangle. Though having its drawbacks, this approach is often used in designing an integer-programming model for site configuration and selection [e.g. Li (1990), Minor and Jacobs (1994), McGarigal and Marks (1995), Williams and ReVelle (1996), and Cova and Church (2000)]. Besides this measure, a number of other direct measures have also been proposed, for example, area-perimeter measures, reference shape, and dispersion of elements within an area. The most popular area-perimeter measure is the Iso-Perimeter Quotient (IPQ) by Osserman (1978). The IPQ is defined as the area (A) times four π divided by perimeter (p) squared ($\frac{4\pi \mathrm{A}}{P^2}$). The second category is to compare the area of a shape to the area of a reference shape, such as the smallest circle that circumscribes the shape (Kim and Anderson 1984). This measure is not scale-invariant and is not additive; therefore, it is not suitable for a regionalization problem in which basic units need to be successively aggregated. The third category is to measure compactness by assessing the dispersion of elements on a shape. The most popular measure of this kind is the moment of inertia approach and the compactness of a shape is defined as $\frac{\mathrm{A}^2}{2\pi I}$, where I is the area moment of inertia of the shape (Li et al. 2013a). These methods, although better than the perimeter approach, are rarely used in optimization models, because they are non-linear and cannot be solved using a integer-linear optimization solver.

This situation has led researchers to implement heuristics, such as TABU, Simulated Annealing (SA), Greedy, and GRASP (Greedy Randomized Adaptive Search Procedure), to solve this difficult problem.

Among popular heuristic algorithms, the greedy algorithm (Black 2005) is often used at the region-construction phase to generate initial region partition plans efficiently, whereas TABU, SA, and GRASP are mostly used at the local-search phase of a regionalization procedure. This phase starts from some initial feasible solution, and moves towards a better solution by reassigning selected atomic polygonal units sitting on the boundary of each region.

The strategy of a greedy algorithm is to always make the best (the optimal, but myopic) choice at each stage with the anticipation that a global optimal solution could be obtained when construction of the region is complete. Although fast, a greedy algorithm often yields a poor local-optimal solution. Examples that fall in this category include Church's PGP (patch-growing process) algorithm for identifying the optimum habitat patch for San Joaquin kit fox (Church et al. 2003), and Guo's contiguity-constrained hierarchical clustering algorithm for clustering electoral votes for Bush in the 2004 Presidential Election (Guo, 2004).

A GRASP algorithm begins with some initial feasible region partition plan and finds improved solutions through local tuning. The local tuning often adopts edge-swapping or edge-reassignment strategies. Edge swapping switches two atomic objects standing on the boundaries of two neighboring regions, and edge-reassignment involves the reassignment of one atomic object on the edge of a region to its neighboring region, if these moves yield an improvement in the solution. The selection of which atomic objects to move is not purely greedy, which always takes the best candidate move; instead, it may be based on randomly selecting a candidate and allowing it to move to an adjacent region if doing so leads to a better solution. Therefore, GRASP can also be considered as a semi-greedy heuristic. GRASP was first introduced by Feo and Resende (1995), and then received widespread adoption in regionalization problems, such as commercial territory design (Rios-Mercado and Fernandez 2009; Rios-Mercado and Salazar-Acosta 2011; Carno-Belmán et al. 2012), districting design (González-Ramírez et al. 2011), and linking habitats for multiple species (Brás et al. 2012).

In GRASP, non-improving moves are not allowed. A TABU search (Glover 1989; Glover 1990), in comparison, allows non-improving moves to move away from local optima. It adopts three strategies. First, a TABU list is introduced to prevent cycles by forbidding reverse moves. This list is dynamic, and after a number of iterations (called TABU tenure) a move is taken out of the list and is then reconsidered as a candidate move again. Second, TABU may force a sequence of up to *k* non-improving moves. Third, TABU search always seeks to make the best greedy move – the move that presents the best improvement or causes the least damage. A TABU search stops when all possible moves are either infeasible (no more uphill moves are allowed and the allowed number of non-improving moves has been reached) or all feasible moves are taboo/banned. A taboo list usually contains elements of solutions that have been recently explored, but it could also involve a list of rules to force the search towards promising areas in the search space, or rules that prevent searching in the part of solution space that contains a known local optima. TABU has found extensive use in regionalization problems, such as ground water parameter zonation (Tung and Chou, 2002), political redistricting (Bozkaya et al. 2003), and Duque et al.'s max-*p*-regions problem (Duque et al. 2012).

Simulated Annealing (SA) is another heuristic allowing non-improving moves when tuning the results. Different from TABU, the non-improving moves are accepted probabilistically. A SA algorithm simulates the process of annealing in metallurgy that involves both a heating and a controlled cooling process. Methodologically, the cooling process starts at a temperature T_0, and descends slowly at a rate of $\propto$ after each round of local adjustment ($\propto$ is always chosen as 0.99 or a higher value such as 0.998). Therefore, the temperature at round k is given by $T = T_0 \propto^{k-1}$. According to SA, if the value of Boltzman equation $e^{\Delta Obj/T}$ is larger than a random number at round k, then a non-improving move will be made. ΔObj is the change of objective value after making a specific candidate move. The whole process will cease when T reaches a predefined value, the temperature at which virtually no nonimproving moves are accepted. Previous work using SA in regionalization include those of Ricca and Simeone (2008) and Duque et al. (2012), and Li et al.'s MERGE algorithm (2014). Specifically, in MERGE, rather than selecting a feasible plan at random, the plan that brings the most increase (or least decrease) in the objective function value of a randomly selected region will be evaluated in the SA procedure. A new stop condition was also added to common SA to intelligently decide when to end the search procedure as well as to keep the best solution found. It also maintains a list of moves that are TABU, allowing the algorithm to exploit a larger search space to obtain better solutions. The adoption of these new strategies ensures better performance of MERGE than common SA and TABU algorithms. In the next section, we will introduce a formulation of the *p*-compact-regions problem that addresses the needs associated with the application of the RELU-TRAN model.

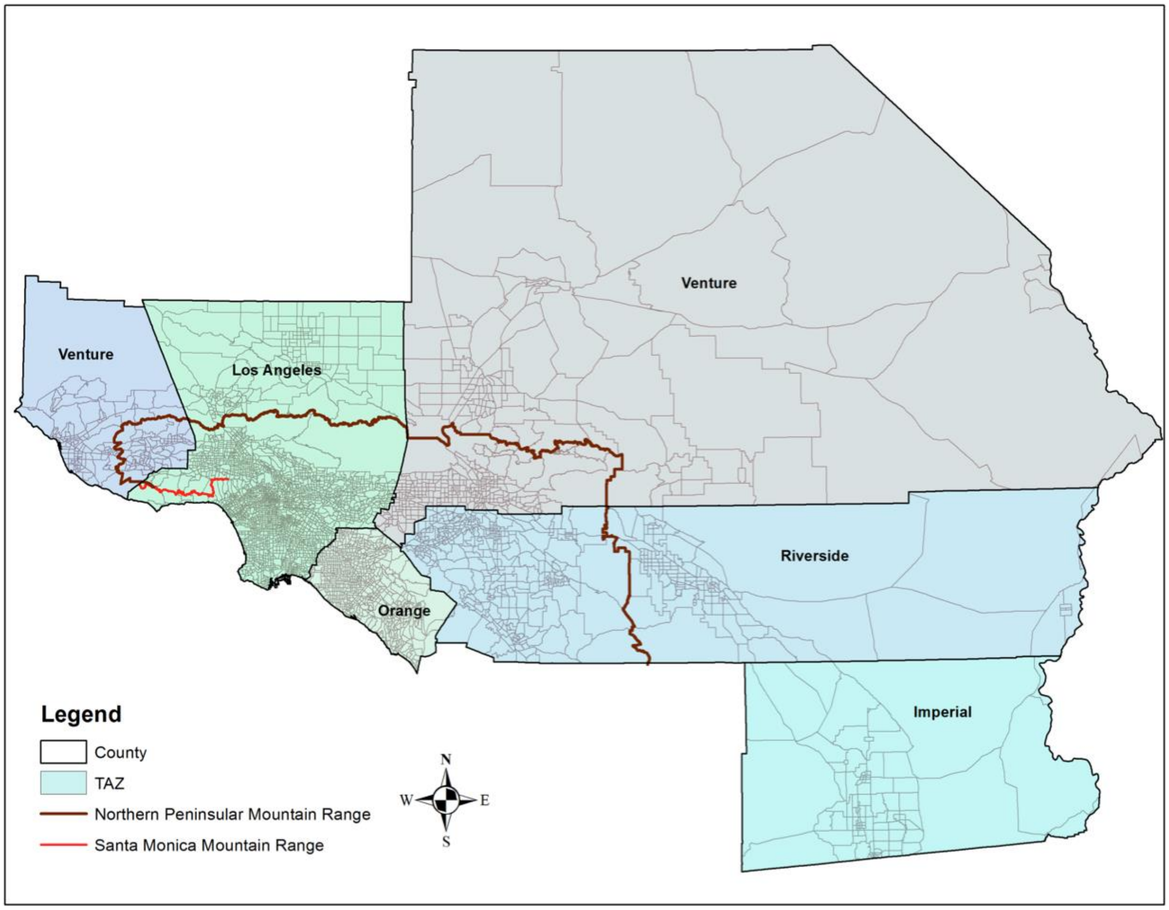


Figure 1. Case study for the *p*-compact-regions problem: grouping more than 4,000 TAZs in six counties of Southern California

3. Problem statement and model formalization

In this section, we introduce the formulation of a real-world *p*-compact-regions problem. Our goal is to aggregate n $(n = 4{,}109)$ TAZs in six counties (Los Angeles, Riverside, San Bernardino, Orange, Imperial, and Ventura, as Fig. 1 shows) of Southern California, U.S., into approximately p $(p = 100)$ model zones, the maximal number considered feasible for the applied urban economic model of RELU-TRAN. Conceptually, we seek to maximize overall *compactness* of the model zones and at the same time preserve the *spatial contiguity* of each zone. A number of linear *physiographic features* have been defined, aligned along major mountain-range barriers, and are used to constrain the placement of model zone boundaries so as to not cross these features. We also constrain model zones from crossing county boundaries. A constraint on traffic flow is also defined, by constraining the solutions such that intra-zonal traffic is less than or equal to a defined proportion of total traffic entering or leaving the region. The traffic flow is measured by the total number of zone-to-zone trips in multiple modes, such as drive-alone trips, shared-ride trips, etc. We also identify a set of economic *subcenters* (large shopping centers and concentrations of employment), and constrain the solution so that subcenters are not split between zones. To formulate this problem, consider the following parameters:

i, j: Index of units that are either subcenters (a small aggregation of TAZs) or independent TAZs (TAZs not part of a subcenter), $i, j \in [1, n]$

u: Index of zones, $u \in [1, p]$

v: Index of known subcenters

t_{ij}: Traffic between unit i and unit j ;

Cty_i: The county that unit i belongs to;

$Tf_i = \sum_{j=1}^{n}(t_{ij} + t_{ji})$: The traffic between unit i and all other units;

$PS = \{(i, j)|$ unit i and unit j are separated by a physiographic boundary$\}$;

$S_v = \{i|$ unit i is a unit of subcenter v and not an independent TAZ$\}$;

$SM = \{i|$ unit i is a member of a subcenter$\}$ where $M = \cup S_v$;

The decision variables are:

$$X_{iu} = \begin{cases} 1, \text{if unit } i \text{ is assigned to zone } u \\ 0, \text{if not} \end{cases}$$

$$T_{iju} = \begin{cases} 1, \text{if unit } i \text{ and } j \text{ are assigned to zone } u \\ 0, \text{if not} \end{cases}$$

Using this notation we can formulate the model as follows:

Maximize:

$$\sum_{u=1}^{p} Compactness(u) \tag{1}$$

Subject to:

$$\sum_{u=1}^{p} X_{iu} = 1, \forall i = 1,2, \dots n \tag{2}$$

$$T_{iju} \leq X_{iu} \quad \forall i, j = 1,2, \dots n; \ \forall u = 1, \dots, p \tag{3}$$

$$T_{iju} \leq X_{ju} \quad \forall i, j = 1,2, \dots n; \ \forall u = 1, \dots, p \tag{4}$$

$$X_{iu} + X_{ju} \leq 1 + T_{iju}, \forall i, j = 1, \dots, n; \ \forall u = 1, \dots, p \tag{5}$$

$$X_{iu} + X_{ju} \leq 1, \forall u = 1, \dots p, \text{ and for } \forall (i, j) \in PS \tag{6}$$

$$X_{iu} + X_{ju} \leq 1, \forall i, j = 1, \ldots, n; \forall u = 1, \ldots p; where\ cty_i \neq cty_j \quad (7)$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} t_{ij}\, T_{iju} \leq \theta \sum_{i=1}^{n} Tf_i X_{iu}, \forall u = 1, \ldots, p \quad (8)$$

$$X_{iu} = X_{ju}, \forall v, \forall i, j \in S_v, \text{ and } \forall u = 1, \ldots, p \quad (9)$$

$$X_{iu} + X_{ju} \leq 1, \forall i, j \in SM, \forall u = 1, \ldots, p \text{ when } i \in S_v\ and\ j \notin S_v \quad (10)$$

constraints to ensure contiguity (11)

The objective function (1) coincides with the definition of a general model of the p-compact-regions problem, aiming to maximize the overall compactness of all model zones. In this case, the total number of atomic units, or TAZs, is 4,109 and the total number of regions, or model zones, is $p = 100$. When X_{iu}=1, it represents the assignment of unit/TAZ i to region u. Constraint (2) ensures that each TAZ is assigned to exactly one region. Constraints (3)-(5) are used to define the value of T_{iju}. $T_{iju} = 1$ when both units i and j are assigned to the same region u. Constraints 3 and 4 ensure that T_{iju} cannot equal one unless both units i and j assign to region u. Constraint (5) forces T_{iju} to equal one when that is the case. Condition (6) is used to prevent the situation in which unit i and unit j are assigned to the same region u when they are separated by a physiographic feature. Constraint (7) keeps two units from different counties from assigning to the same region. Constraint (8) sums all traffic between every single pair of atomic units within region u, that is, intra-zonal traffic, and stipulates that this intra-zonal traffic must be less than or equal to a fraction θ of this zone's total traffic, obtained by summing up all traffic that occurs between any pair of atomic units within region u and all atomic units (including itself) within the study area. This constraint keeps the property between zones (TAZ and model zone) on different geographical scales consistent, because the design principle of a TAZ is also to capture as many inter-zonal trips as possible. In another word, loss of the intra-zonal trips should be prevented (Miller and Shaw, 2011). Constraints (9) ensure that TAZs that belong to the same subcenter must be assigned to the same region. Constraints (10) prevent two TAZs that are members of the different subcenters from assigning to the same region. In addition, spatial contiguity constraints (11) within each model zone must be satisfied as well. Examples of spatial contiguity constraints are described in Duque et al. (2012). One of the forms discussed in Duque et al. (2012) would need to be appended to the above formulation to make it complete. The size of the problem application discussed here falls outside the range of problem sizes that can be solved to optimality, when the compactness measure is structured in an integer-linear form, such as minimizing the length of the perimeter (outside boundary) of each region. Unfortunately, one of the best measures of compactness is based upon the moment of inertia, which is a non-linear formula. Consequently, the general form of the *p*-compact regions problem, involves a non-linear objective as well. Both size and non-linear features force the development of a heuristic for this problem.

4. Methodological framework for the model

In order to find a near-optimal solution to the *p*-compact-regions problem, we propose a methodological framework based on the combined use of randomized greedy for region generation and an edge-reassignment-based local search to further tune and improve the results. The Region-construction phase is designed to assign all available basic units into the desired number of regions. As this assignment considers all basic units in the study area, it can be considered as a global strategy. Whereas, the local search phase designed is to fine tune the generated regions by reassign the TAZs on the edge of a model zone when such moves improve the objective. Before diving into the detailed discussion about the algorithm, we first introduce some important data structures.

Figure 2 illustrates three classes to describe an important set of data: model zone, TAZ and a region growth/change plan. In a model zone class, properties of a model zone, such as its area, perimeter, moment of inertia (used for computing compactness), and the compactness value of a model zone are defined. The sum of the compactness values of all model zones equals the objective function value for our problem. Other properties, such as intra-zonal traffic are recorded as well. This value is computed by knowing the list of TAZs inside of the current model zone. Therefore, the TAZs that a model zone contains are also an attribute of the model zone class. When all model zones finish growing, namely, when all TAZs are assigned, an overall model zone partition plan is generated by collecting the list of TAZs that belong to each model zone. Another class is the TAZ class, which records the area, perimeter, its moment of inertia and neighbor TAZs of a TAZ. Two TAZs are considered as neighbors if they share a common arc. These attributes of a TAZ are used for computing the objective function value of the model zone to which it belongs. Though sharing common attributes, the TAZ class and the model zone class are different in essence. The information in the TAZ class is static throughout the regionalization process; whereas the information of a model zone changes as a region grows in size or changes. A third class is a growth or change plan for a model zone. This class is composed of three elements: a TAZ to add to or remove from a model zone; the change of the objective function value (the change of compactness value of this model zone) brought from this change; and the properties (listed in the model zone class) of the model zone after this change is made. In the region growth phase, such plan always involves adding another TAZ; whereas in the local search phase, such plan may be involve removing a TAZ from a model zone or adding a TAZ. These data structures ensure the efficient execution of the regionalization algorithm by 'memorizing' the status of each model zone at each phase and the potential new status of a model zone if it is selected for the change (e.g. growth).

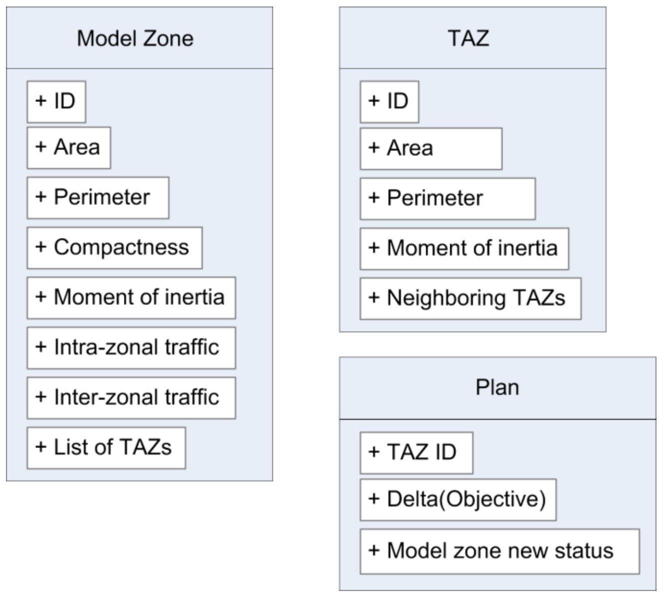


Figure 2. Class design for model zone, TAZ and a region growth plan.

Due to the nature of the objective and constraints, there is a need to tailor and address the following issues when implementing the region-growing procedure. First, it was decided that major employment subcenters, usually a small group of TAZs should be assigned to only one model zone as well as ensure that each model zone contains no more than one of these subcenters. Therefore, these subcenters are ideal seeds for growing the regions. We added to these a select set of other TAZs for seeds (the selection procedure is discussed in detail in Section 5.1), so that there were a total of 100 individual seeds. The purpose of the algorithm is to assign all of the other TAZs in our study area to one of the 100 zones with the strict restriction that each TAZ can only be assigned to one model zone.

Second, at each time only one TAZ will be assigned to a model zone, meaning that rather than growing simultaneously, only one model zone grows at a time. The TAZ that is assigned must lead to an improvement in the objective function (or harm it the least) that is at or close to the greatest possible improvement (or least harm) over all possible growing plans for a model zone. We choose to not always assign the TAZ with the best impact on the solution because this strategy may not yield a good final solution. Rather than using a strict greedy algorithm, a probabilistic selection strategy is introduced to randomly select one of the best *N* choices to grow a model zone (Li et al. 2014). Using this strategy, a different model zone partition plan will be generated after each complete execution of the region growth procedure. This allows us to generate a large number of different, but good solutions with which we can apply an improvement process. Essentially, the process of growing regions involves two steps: 1) dealing, and 2) randomized greedy growth. The randomized greedy process will be described next.

Figure 3 demonstrates the pseudo code of the randomized greedy algorithm with time complexity O(np), where *n* is the number of TAZs and *p* is the number of model zones. The first step is to explore and identify all possible plans to grow all models zones. Then for each model zone, the best *N* plans are saved. A best plan for a model zone is one which brings the most increase in its compactness by adding an unassigned TAZ. Next one plan is randomly selected from among the *N* plans for each model zone such that each zone now has a candidate plan to grow. .After this step, these candidate plans will be ranked, and the overall best plan (adding TAZ i to model zone x) will be executed. This includes updating model zone *x*'s properties, including area, perimeter, moment of inertia and adding TAZ *i* to the list of TAZs that model zone *x* contains. At the same time, TAZ *i* will be labeled as 'taken' to avoid duplicated assignment of this TAZ. After a model zone grows by an added TAZ, a new candidate plan needs to be identified for it. Any other models zones whose candidate plans involved adding the recently assigned TAZ *i* will need to be reexamined to find another candidate plan as a replacement. The examination process is the same as what it does in the first step. Because candidate plans are saved for other unaffected model zones, there is to need to reexamine those model zones. This strategy greatly improves the efficiency of the algorithm. The randomization in this process occurs in the selection of the best of the *N* plans for each zone. This strategy enlarges the effective search space even when the value of *N* is small ($N>1$). Note, when $N=1$, this process would condense to the classic greedy heuristic.

```
for each model zone k:
    Explore all possible plans to grow the model zone;
    Find and save the top N best plans that bring the most increase in the model zone's compactness;
    Randomly select one of these N plans as the candidate plan to grow model zone k;
end for
while (if there remain TAZs not assigned to a model zone)
    Rank through all candidate plans, and select the model zone x that has the best plan:
        <TAZ i, delta(objective), new status for model zone x >;
    Make model zone x to grow, this includes:
        (1) update model zone x's properties by those defined in the plan;
        (2) add TAZ i to the list of TAZs belonging to model zone x;
        (3) label TAZ i's status as 'taken'.
    for model zone x and any other model zone that list TAZ i in its candidate plan:
        Explore all possible plans to grow the model zone;
        Find the top N best plans that bring the most increase in the model zone's compactness;
        Randomly select one of these N plans as the candidate plan to grow this model zone;
    end for
end while
```

Figure 3. Pseudocode for region growth algorithm

After executing the randomized greedy algorithm, an initial model zone partition plan is generated. To further improve the quality of the results, we employ after all TAZs have been assigned, an edge-based reassignment process. This edge-reassignment-based local-search module examines each TAZ that lies at the edge of a model zone, and evaluates moving it from its current model zone to the one across the edge/boundary, to seek a better solution with the requirement that no constraints are violated. The procedure will cease when no better solutions can be obtained by edge reassignment. The process of edge reassignment can be controlled by a number of strategies, like simulated annealing and TABU search.

Figure 4 shows the system diagram of the solution framework. Three modules on the left of this figure stand for phases of seed selection, region construction and local search from top to bottom. The algorithm for region construction is greedy-based, with two strategies being used: "dealing" and randomization. The role of dealing is to make each region grow to a viable size by assigning (dealing) a fixed number of TAZs to the seeded regions. This dealing process avoids the generation of regions with weird shapes. It also prevents regions being generated that are very small but compact. In a compactness-driven regionalization problem, once a small and compact region is formed, adding any adjacent TAZ to it will decrease its compactness. Therefore, it will have very little chance to grow because this zone's candidate plan has a negative value rather than a positive value. So the neighboring TAZs of the small and compact regions have to be taken by neighboring regions, resulting in regions of weird shapes, such as concave shapes. By adapting the dealing strategy, each region is forced to grow to at least a realistic starting size. After the dealing phase, the algorithm moves towards finishing the assignment of all available TAZs using the randomized greedy selection method, discussed above. As randomization increases the solution space after running the region-construction algorithm a number of times, a set of initial region partition plans can be constructed. These plans will be fed into a highly extendable local-search module, in which we can plug in and test the performance of several popular meta-heuristics, including GRASP, TABU, and SA. When testing SA, Li et al.'s MERGE (Li et al. 2014) algorithm is used, as it is an SA-inspired algorithm and is proven to perform better than the basic SA algorithm.

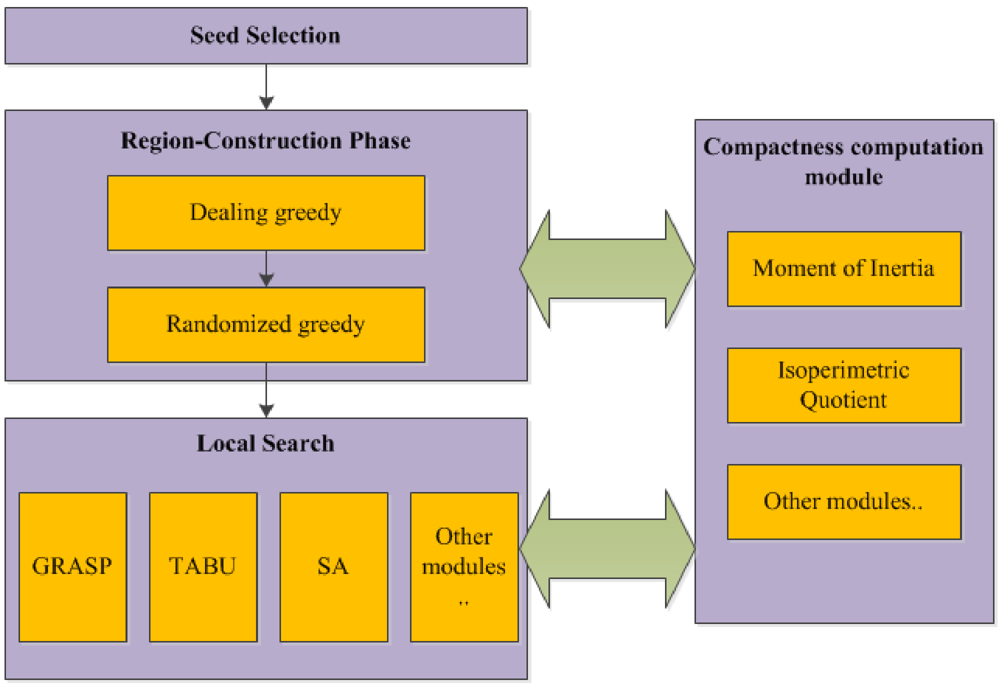


Figure 4. System diagram

At each stage of region construction and local search, the algorithm module communicates intensively with the compactness computation module, as each time a region's shape changes, its shape compactness needs to be re-calculated. The compactness computation module is also highly extendable, allowing the integration of various shape compactness measures, such as the Normalized Moment of Inertia (NMI) (Massam and Goodchild 1971; Li et al. 2013), the commonly used Isoperimetric Quotient (IPQ; Osserman 1978), or some indirect measure.

During the region-construction process, the TAZ added to a model zone is selected from a pool of all unassigned TAZs adjacent to the model zone. Therefore, contiguity of model zones is preserved without the need for a specific strategy. In the local-search phase, however, removing one TAZ from a model zone may potentially violate the contiguity constraint by leaving a residual zone that is split into two unconnected parts. To address this issue, we implement an efficient algorithm to identify the contiguity of a region. This algorithm relies on a dictionary recording common boundaries (arcs) between atomic units within a region, indexed by IDs, and another dictionary storing the pointers used in the contiguity checking process. The ID of a unit is the key and the ID of a neighbor to which this unit points to is its value. The algorithm operates as follows. For each basic unit set an initial pointer to zero. Read the first common boundary (arc) record. Suppose it indicates a common boundary between k and l. If neither k nor l's pointer has been set, set the pointer of the greater of $\{k, l\}$ to point to the lesser of $\{k, l\}$. If the pointer of k is already set, follow the pointers until they terminate, say at unit i, and make l point to i. Similarly if the pointer of l is already set, follow its pointers and set the pointer of k. If pointers are already set for both k and l then go to the next record. Continue to process all records. When the records are finished, scan the pointers (the value field in the dictionary). If all pointers have a single non-zero value the region is contiguous. If pointers have two or more non-zero values there are islands. This algorithm can perform the contiguity check rapidly as it tests contiguity with a single pass of the boundary records.

5. Implementation of the computational model

In this section, we discuss in detail how the objective and constraints of the model are computed quantitatively.

5.1 Selection of seeds

The seeds are the starting points for growing model zones. As noted earlier, we defined a set of subcenters, which are areas where the employment concentration is relatively high (McDonald 1987). In this paper, these employment subcenters are contiguous (not necessarily compact) areas composed of TAZs, as shown in Figures 5(a) and (b). Based on Giuliano and Small's procedure (1991), 51 subcenters were identified in our study area, including 19 in Los Angeles, 11 in Riverside, 10 in Orange, 6 in San Bernardino, 4 in Imperial, and 1 in Ventura Counties. One subcenter cannot be assigned to more than one model zone, and one model zone cannot contain more than one subcenter. Therefore, the subcenter itself can be a natural seed. Using subcenters as seeds to grow model zones results in the following advantages: first, Freeways and highways are designed to join subcenters. If one divides a metropolitan area into zones, one may choose to employ an aggregated representation of the freework/highway network. The aggregated freeway/highway network will better resemble (and propagate fewer aggregation errors) the actual network if the seeds are subcenters. Second, for transit-oriented development, transit networks tend to link subcenters. Employing zoning based on subcenters makes it easier to analyze proposed transit lines using an aggregated network. Third, people tend to conceptualize urban space in terms of subcenters. Thus, a zoning system based on subcenters will tend to accord with intuition. Fourth, in central place theory, economic activity is organized spatially with reference to the central places. Subcenters are central places.

As we need 100 model zones in total, another 49 TAZs are selected randomly or manually as the supplemental seeds to the existing subcenter seeds. To select these supplemental seed TAZs we first identified junctions in the major-highway network, which tend to be areas of high population density. This selection was based upon the fact that such areas represented places of attraction for trips, and the intent of econometric model was to model land use, traffic, and housing and commercial growth. We also selected some arbitrarily TAZs in order to augment the subcenters and major-highway junctions. The principle for seed selection was to place seeds as dispersed as possible and denser highly populated areas received more seeds. With the exception of the subcenter seeds, the remaining seed TAZs served only to initialize the growth of model zones. During the edge-reassigning process, these TAZs can be swapped out of the original model zone and therefore will not have a great effect on the final result.

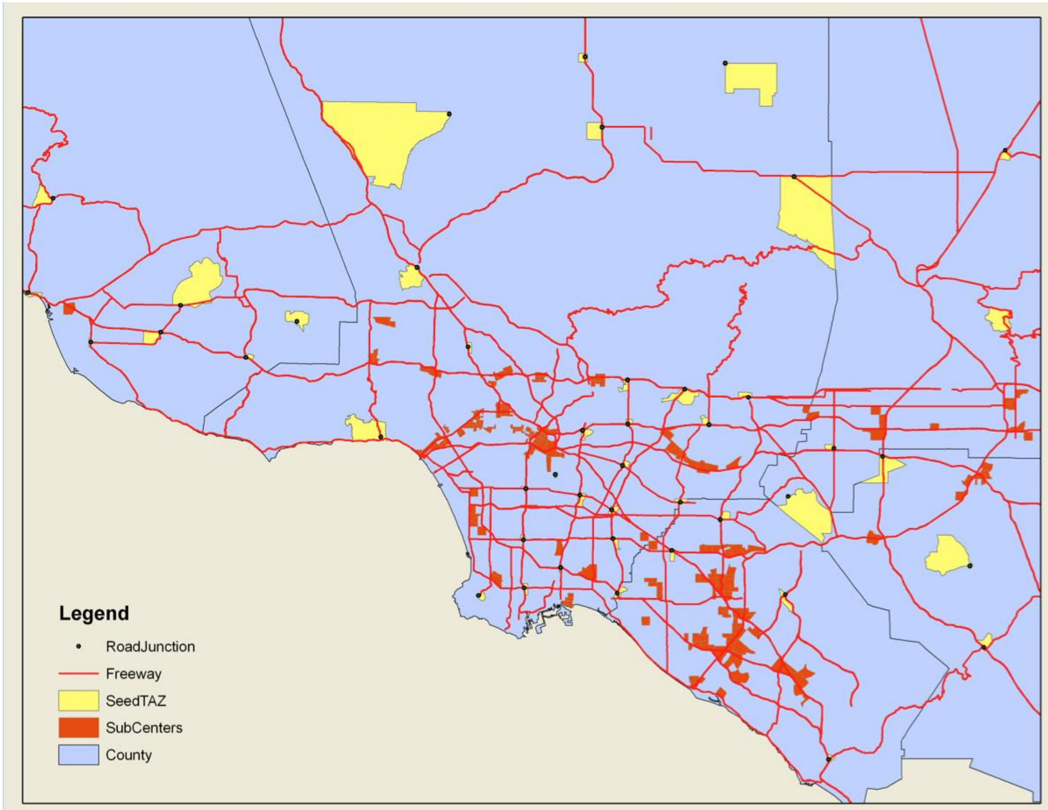


Figure 5(a). Seed TAZs in Ventura, Los Angeles, and Orange Counties.

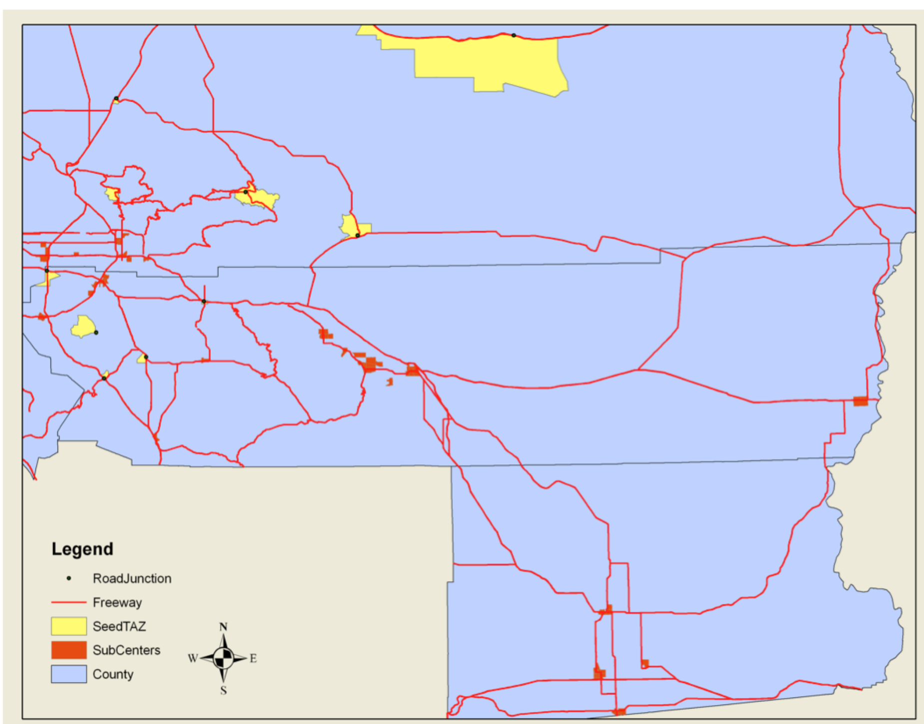


Figure 5(b). Seed TAZs in San Bernardino, Riverside, and Imperial Counties.

5.2 Compactness measure

We implemented the compactness measure using the Normalized Moment of Inertia (NMI) as described in Li et al. (2013a) and the commonly used Isoperimetric Quotient (IPQ) approach. The computation of compactness in a regionalization process using either approach includes three parts: (1) compactness index of 4,109 TAZ shapes, computed about the shape's centroids; (2) the change in compactness of a model zone when a new TAZ is added during the region-construction phase; and (3) the change of this property when a TAZ is detached during edge reassignment at the local-search phase.

$$NMI_k = \frac{A_k^2}{2\pi I_k} \tag{12}$$

$$IPQ_k = \frac{4\pi A_k}{P_k^2} \tag{13}$$

$$I_k = \int d^2 da \tag{14}$$

Equations (12) and (13) demonstrate how to compute the compactness index of a single shape using NMI and IPQ. A_k is the area of a single TAZ or a region; k is its index; P_k refers to the length of its perimeter; I_k is the second moment of a single TAZ of region k about an axis perpendicular to the region's surface and passing through its centroid. Mathematically, I_k is equal to the integral of the squared distance from any infinite small area to a region's centroid, as shown in Equation (14). Detailed calculation of I_k can be found in Li et al. (2013a). Both of these two measures have the same value range of (0,1]. A higher value indicates a more compact region and a circle receives the highest compactness index 1. By calculating the area and second moment of a TAZ, the compactness index NMI can be obtained. Similarly, the IPQ can be obtained for a TAZ once its area and perimeter are known. When a region grows, the area, perimeter, and second moment all change and can be computed by Equations (15)-(17):

$$A_{Z'_k} = A_{Z_k} + A_{TAZ_i} \tag{15}$$

$$P_{Z'_k} = P_{Z_k} + P_{TAZ_i} - \sum_j len\left(arc_{TAZ_j TAZ_i}\right) \tag{16}$$

$$I_{Z'_k} = I_{Z_k} + I_{TAZ_i} + d^2_{G(Z_k)G(Z'_k)} A_{Z_k} + d^2_{G(TAZ_i)G(Z'_k)} A_{TAZ_i} \tag{17}$$

, where Z'_k is the new region after adding a TAZ i, $len(arc_{TAZ_j TAZ_i})$ is the length of the arc shared by TAZ i and any TAZ_j belonging to Z_k. G is the centroid of a TAZ for a region, and d is the Euclidean distance between the two centroids (region and TAZ). As both the computation of NMI and IPQ are additive, a new data structure is defined to carry the parameters including area, perimeter, centroid, and second moment of both single TAZs and partial regions to calculate NMI and IPQ of a region as it changes. Therefore, once a region's shape changes, there is no need to compute these values from scratch, but only the need to apply Equations (15)-(17).

5.3 County boundaries and physiographic features

One of the constraints in this optimization model is that a model zone cannot cross county boundaries or physiographic barriers. As shown in Fig. 1, TAZs are partitioned within counties, and therefore no single TAZ crosses a county boundary. To force model zones not to cross county boundaries, a topological analysis of (TAZ, County) containment was conducted and a hash table was established with the ID of the TAZ as the key and the county that it belongs to as the value. Physiographic features act as natural boundaries for model zones as well, since it makes little sense for a local zone gained for the purposes of modeling land use, economic, and transportation planning to have a mountain range running through it. In this project, two mountain ranges in Southern California are identified as physiographic barriers: the Peninsular Mountain Range and the Santa Monica Mountains. The Peninsular Mountain range crosses four counties and forms closed boundaries with county boundaries, dividing the four counties into sub-counties. Since neither the county nor physiographic boundary can be crossed, the values in the hash table can be defined as sub-county codes in place of the original county codes. When a candidate TAZ is selected to be added to a model zone, the sub-county code will be fetched from the hash-table lookup. If this TAZ is not in the same sub-county as all the other TAZs in the model zone, this TAZ will not be considered for adding to that zone.

The Santa Monica Mountain barrier requires a different approach since it does not form a closed boundary with county boundaries (specifically Los Angeles County). To incorporate this physiographic barrier into the zoning process, conflict groups were constructed of TAZs that cannot be assigned to the same model zone because they fall on opposite sides of the physiographic barrier. Each time a TAZ is considered for assignment to a model zone, the program will check whether it has a conflict with any TAZ that has already been included in the model zone. The above strategy guarantees that both county boundary and physiographic feature constraints are satisfied.

5.4 Modeling zonal traffic

To restrict the size of each model zone and to keep the errors in modeling transportation flows on the highway network small, intra-zonal traffic is limited to be no more than a fraction of the inter-zonal traffic of a region.. Define Z_k as the set of TAZs comprising model zone k. Let t_{ij} denote the traffic observed between TAZ *i* and TAZ *j* (we discuss the nature and source of such data below). Finally, let $IntraT_{Z_k}$ represent the intra-zonal traffic in model zone k, and $InterT_{Z_k}$ represent the inter-zonal traffic into and out of zone k:

$$IntraT_{Z_k} = \sum_{i \in Z_k} \sum_{j \in Z_k} t_{ij} \tag{17}$$

$$InterT_{Z_k} = \sum_{i \in Z_k} \sum_{j \notin Z_k} (t_{ij} + t_{ji}) \tag{18}$$

Our goal is to limit the ratio between intra-zonal trips and the sum of intra- and inter-zonal trips to be less than or equal to a threshold θ:

$$\frac{IntraT_{Z_k}}{IntraT_{Z_k} + InterT_{Z_k}} < \theta \tag{19}$$

The assumed value of θ is 0.1. When all model zones reach this intra-zonal to inter-zonal traffic limit but there are still some TAZs unassigned, θ is automatically increased by 0.05. This operation is repeated until all TAZs are assigned. To calculate variables $IntraT_{Z_k}$ and $InterT_{Z_k}$, a TAZ-by-TAZ origin-destination matrix containing the number of trips between each pair of TAZs was obtained from the Southern California Association of Governments.

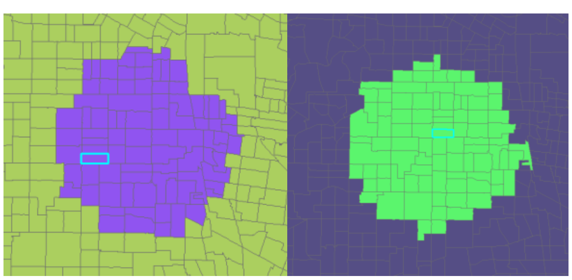
6. Experiments

The above section discussed the computational issues of implementing the regionalization process. In the next section, we describe several experiments to explore variations of these methods and assess their effectiveness in solving this applied p-compact-regions problem.

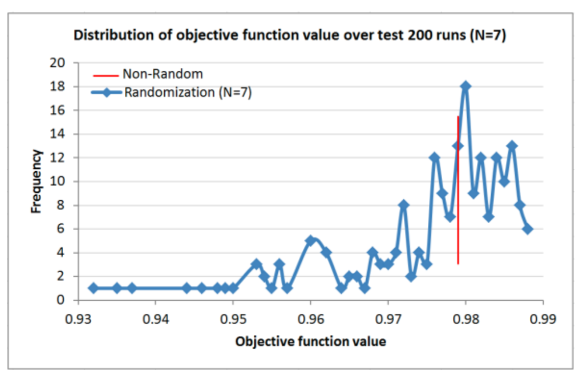
6.1 Impact of Randomization

Randomization is introduced in the model-zone construction phase to generate a larger solution space by randomly selecting one plan from the top *N* best plans for growing a model zone. In this section we investigate appropriate settings for this randomization strategy. Figure 6(a) demonstrates the growth of a region from a seed TAZ in L.A. County (highlighted in blue) using non-randomization (left) and randomization (right; N=5) approaches. Both model zones start to grow from the same seed TAZ, and stops when the intra-zonal traffic reaches the threshold (10% of the total traffic). It can be said that both strategies lead to relatively compact regions, however, the regions grew in different directions: one grew towards the northeast of the seeded TAZ and the other grew almost evenly around the seed TAZ. Figures 6(b)-(d) demonstrate the distribution of the objective (the compactness of a region using the NMI approach) over 200 runs when *N*=7, 5, and 3 respectively. The red vertical line in the figures represents the achieved objective value 0.978 when the non-randomization approach is applied. It can be seen clearly that by using randomization, the majority of the 200 results (over 50%) are better than when non-randomization is used. In addition, using the non-randomization approach, the region will always grow in the same direction and shape as the left figure in Figure 6(a). For a randomization approach, a region could possibly grow in any direction and the right figure in Figure 6(a) shows just one of many possibilities. This increases the flexibility of region generation and is especially helpful when multiple regions are growing simultaneously. Therefore, adopting randomization in the region-construction process certainly performs better than when it is not used.

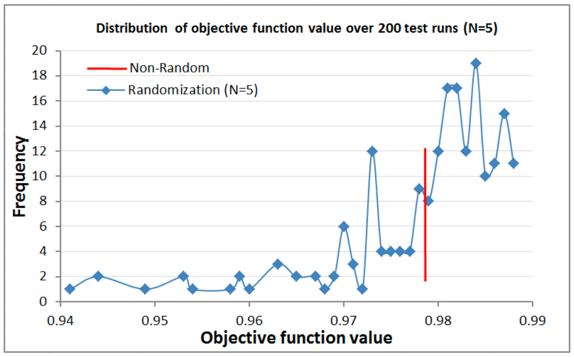
We also examined the optimal setting of *N* in this randomization strategy. The stopping criterion for a run is the same as that set in the experiment shown in Figure 6(a). We set the length of the best moving plans *N* to be 3, 5 and 7 in the experiments, and the results show that when *N*=3, over 75% of the simulations lead to plans better than 0.978, while when *N*=5 or *N*=7, the ratios are only 63% and 56%. Therefore, *N* = 3 is the best setting of *N* and is used in the solution framework algorithm.



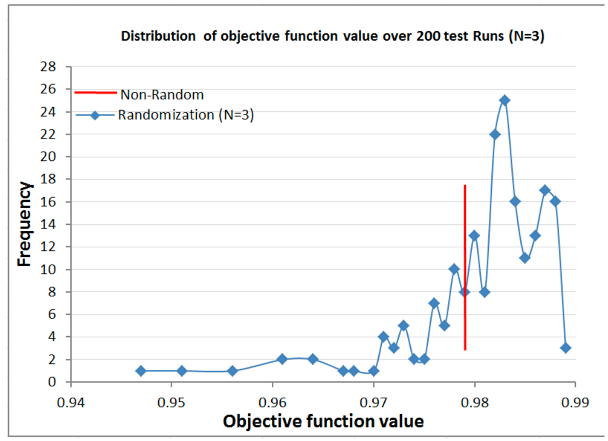
(a)



(b)



(c)



(d)

Figure 6. Distribution of overall objective function values of a test model zone in LA County over 200 runs.

6.2 Performance comparison of meta-heuristics at the local-search phase

The meta-heuristic modules integrated in our solution framework, including GRASP, TABU, and MERGE, were each run 500 times over 500 feasible solutions generated from the randomized region-construction phase to identify the best solution. The *Y* axis shows the overall objective value, which is the sum of the compactness index (using NMI) over 100 model zones. As the range of NMI for a single region is (0,1], the range of objective values equaling the sum of NMI over 100 zones is (0,100]. The *X* axis shows the N^{th} run of a meta-heuristic. A trend line using a moving average is also shown. The red dots/line show the objective value of 80.5 that can be achieved using the non-randomized greedy algorithm. Clearly, better solutions can be obtained when adopting randomization and meta-heuristics including GRASP, TABU, and MERGE. Among these meta-heuristics, MERGE is demonstrated to be the best by producing zones with an average compactness value of 0.89 while satisfying all other constraints.

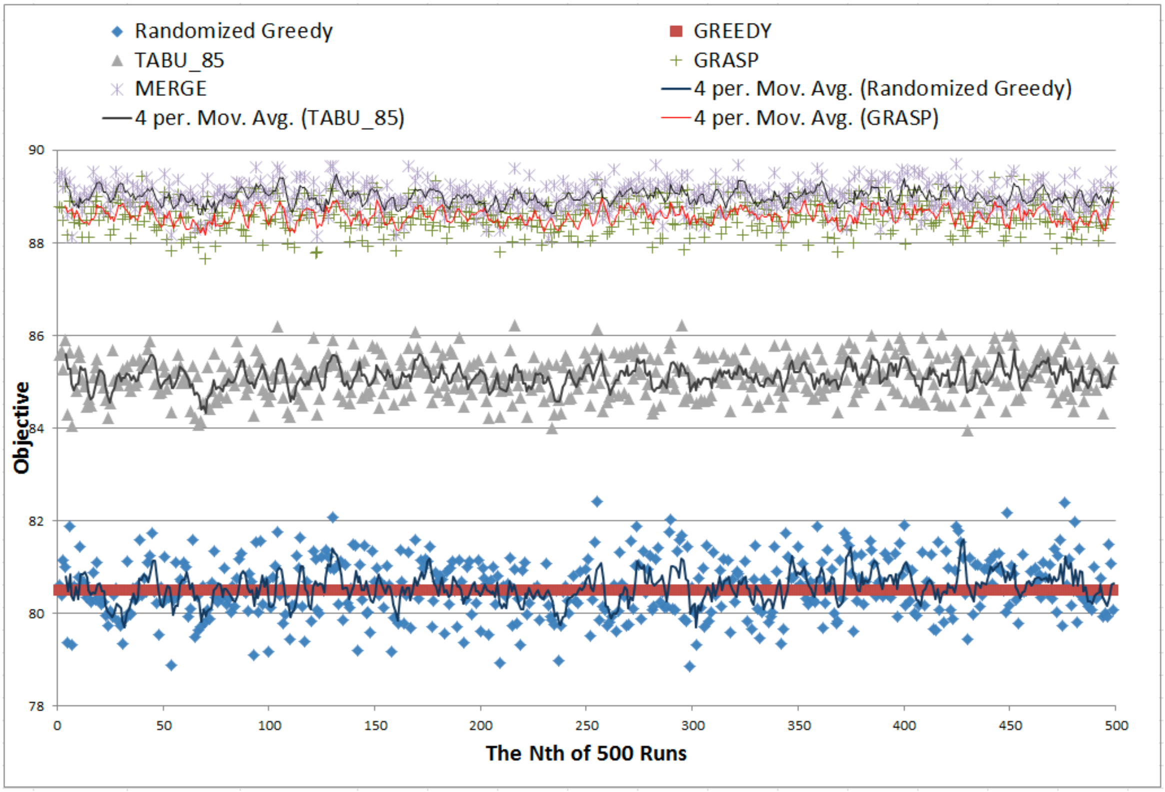


Figure 7. Performance comparison in achieving optimality using metaheuristics. "4 per. Mov. Avg" means the trend line generated by taking the arithmetic mean of subsequences of four data points.

6.3 Comparison of region compactness using NMI and IPQ measures

The best solutions from the 500 runs and the zonation plans using IPQ and NMI as compactness measures are shown in Figure 8 and Figure 9. Intuitively, the model zones produced by the proposed compactness measure NMI are much rounder in shape and insensitive to the detailed form of the edge. The average NMI compactness of model zones constructed using IPQ is 0.711, which is 0.18 less than for zones constructed using NMI. Figure 10 counts the number of regions out of 100 with shape indices falling in each bin. We can see that for the regions generated by the NMI approach, about 70% achieved a compactness of 0.9. However, for those generated by the IPQ approach, only less than 20% had a compactness equal to or greater than 0.9. This experiment verified that the NMI approach substantially outperforms the IPQ approach in a regionalization problem aiming at generating compact regions. This is because, in principle, the IPQ method suffers from the well-known bias introduced by estimating the lengths of real geographic curves from the lengths of their polyline representations (Longley et al. 2010), as opposed to the NMI, which is directly related to the average accessibility of the entire area, rather than the geometry of its perimeter. The contrast between the two measures is most obvious when a roughly circular area that in fact has a long, undulating perimeter several times as long as the perimeter of a circle is compared to a long, thin area. The IPQ will return similar values in these two cases, whereas the NMI will be very different, indicating the marked difference in accessibility and circularity of the two cases. Therefore, the NMI approach was adopted in our solution framework for this *p*-compact-regions problem applied to the Southern California Association of Governments area.

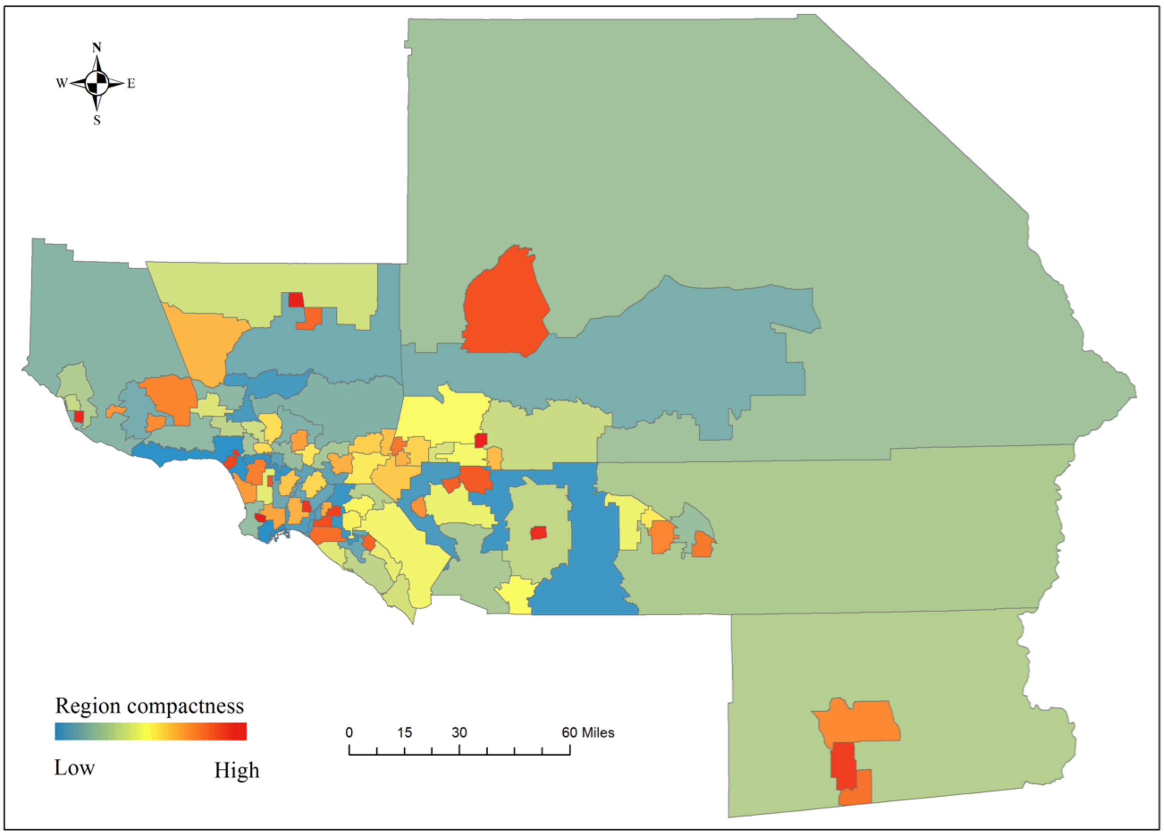


Figure 8. Zoning result using IPQ as the compactness measure

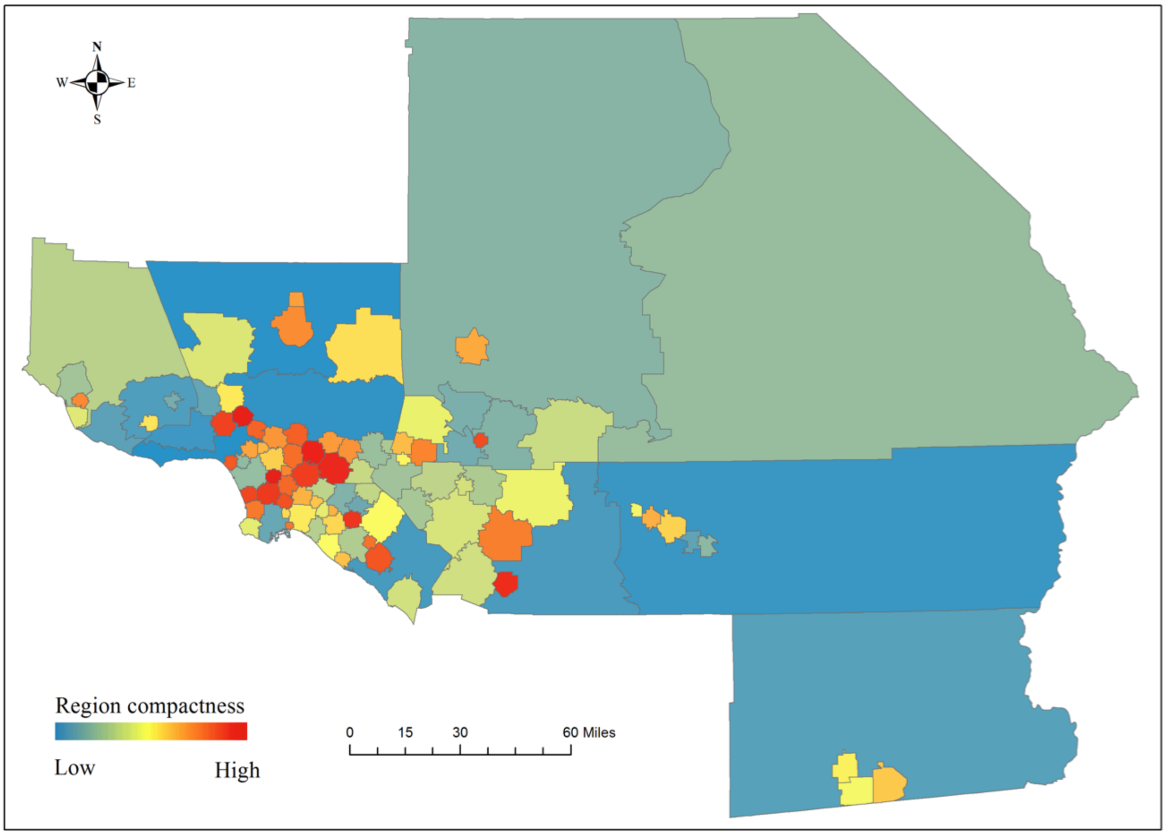


Figure 9. Zoning result using NMI as the compactness measure

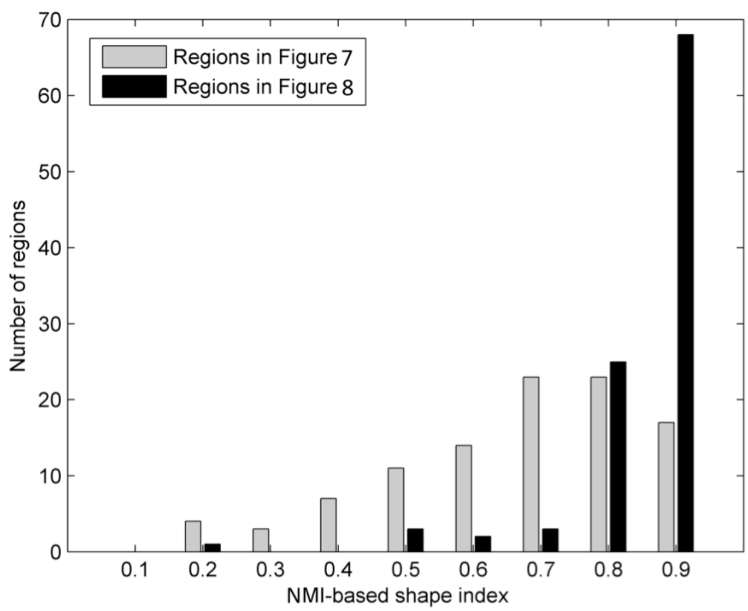


Figure 10. Comparison of compactness distribution of the 100 generated model zones in Figure 8 and Figure 9

7. Discussion and Conclusions

This paper presents a real-world *p*-compact-regions problem and proposes an extendable heuristic framework to solve it to near-optimality. This TAZ zonation problem is substantially different from those described in the literature both conceptually and computationally, because of the large number of building blocks and zones, the complexity of the objective function and constraints, and the innovative approach to compactness. We introduced the model and the computational issues for both the objective and constraints, the design and implementation of the solution framework, and the impact of a number of techniques, including various meta-heuristics, compactness measures, and the quality of the results based upon the degree to which a randomization strategy was used. Through a series of experiments, we concluded that when the number *N* of candidate building blocks maintained during the region construction phase is equal to 3, using MERGE for local search and NMI as the compactness measure yields the best solution in comparison with other techniques.

The RELU-TRAN project is still under development, this generation of model zones being one of the first stages in a multi-year project. At this time the modelers in the research team are quite satisfied with the model zones, finding upon detailed inspection that they satisfy all of the requirements both in principle, in terms of the objectives and constraints, and in practice in terms of the detailed positioning of zone boundaries. The project will now proceed to the next steps of model calibration and application.

In summary, our purpose in this paper was to explore various aspects of the *p*-compact-regions problem, using a real, practical example of substantial size. We make an original contribution in the use of NMI as a compactness measure in a *p*-compact-regions problem and prove its better performance over the commonly used IPQ approach. As the NMI approach is additive, robust towards positioning errors, and is able to handle regions with holes or multi-parts, it can well handle the non-continuous study area, such as those with a lake removed or having islands in the regionalization process. This is a great advantage in using NMI over other compactness measures. This paper also makes a substantial contribution in providing a highly extendable solution framework that allows other researchers to investigate a large, real, non-linear regionalization problem and find practical solutions. We believe that the experience we have gained in this project, and the solutions we have found to many of its issues, may provide useful guidance to others.

Though generally applicable, we believe that any implementation of the *p*-compact-regions problem must respond to its specific circumstances, as applications will vary in the number and nature of constraints, the number and nature of other objectives, the computational issues raised by the size of the problem, and the geography of the application area. For example, when applying the proposed solution framework to other applications or similar application with larger size, the randomization parameter *N* needs to be carefully tuned for a proper sampling of solutions in the solution space. Although the optimization model we formulated here is single-objective tailored to the specific need of the urban economic modeling problem, it will be highly extendable to integrate more objectives. Preliminary experiments have been conducted in using an objective which minimizes intra-zonal dissimilarity instead of maximizing compactness in our solution framework. The results demonstrated that the proposed solution framework is able to solve the problem with much less time and can still achieve the same level of solution quality in comparison to a commercial solver CPLEX. In the future, we will conduct a systematic testing on a multi-objective regionalization problem and propose a more comprehensive solution framework for broader applications. We are also working on integrating the *p*-compact-regions code with the open-source spatial analysis library PySAL (http://pysal.org) to make it openly available to peers in the GIS community.

References

Alsalloum, O. I., & Rand, G. K. (2006). Extensions to emergency vehicle location models. *Computers & Operations Research, 33*, 2725-2743.

Anas, A., & Liu, Y. (2007). A Regional Economy, Land Use, and Transportation Model (RELU-TRAN): Formulation, Algorithm Design and Testing. *Journal of Regional Science, 47*, 415-455.

Angel, S., Parent, J., & Civco, D. L. (2010). Ten compactness properties of circles: measuring shape in geography. *Canadian Geographer-Geographe Canadien, 54*, 441-461.

Black, P. E. (2005). Greedy algorithm. *Dictionary of Algorithms and Data Structures*. U.S. National Institute of Standards and Technology (NIST).

Bozkaya, B., Erkut, E., & Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research, 144*, 12-26.

Brás, R., Cerdeira, J. O., Alagador, D., & Araújo, M. B. (2012). Linking habitats for multiple species. *Environmental Modelling & Software, 40*, 336-339.

Cano-Belmán, J., Ríos-Mercado, R. Z., & Salazar-Aguilar, M. A. (2012). Commercial territory design for a distribution firm with new constructive and destructive heuristics. *International Journal of Computational Intelligence Systems, 5*, 126-147.

Church, R. L., Gerrard, R. A., Gilpin, M., & Stine, P. (2003). Constructing cell-based habitat patches useful in conservation planning. *Annals of the Association of American Geographers, 93*, 814-827.

Church, R. L., Stoms, D. M., & Davis, F. W. (1996). Reserve selection as a maximal covering location problem. *Biological conservation, 76*, 105-112.

Cova, T., & Church, R. (2000). Exploratory spatial optimization in site search: a neighborhood operator approach. *Computers, Environment and Urban Systems, 24*, 401-419.

Cova, T. J., & Church, R. L. (2000). Contiguity constraints for single‐ region site search problems. *Geographical Analysis, 32*, 306-329.

Duque, J. C., Anselin, L., & Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science, 52*, 397-419.

Feo, T. A., & Resende, M. G. C. (1995). Greedy randomized adaptive search procedures. *Journal of global optimization, 6*, 109-133.

Fischer, D. T., & Church, R. L. (2003). Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model. *Forest Science*, *49*, 555-565.

Gendreau, M., Laporte, G., & Semet, F. (2005). The maximal expected coverage relocation problem for emergency vehicles. *Journal of the Operational Research Society*, *57*, 22-28.

Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region. *Regional Science and Urban Economics, 21*, 163-182.

Glover, F. (1987). Tabu search methods in artificial intelligence and operations research. *ORSA Artificial Intelligence*, *1*, 6.

González-Ramírez, R., Smith, N., Askin, R., Miranda, P. A., & Sánchez, J. (2011). A hybrid metaheuristic approach to optimize the districting design of a parcel company. *Journal of applied research and technology, 9*, 19-35.

Guo, D. (2009). Greedy optimization for contiguity-constrained hierarchical clustering. *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on* pp. 591-596). IEEE.

Hess, S. W., & Samuels, S. A. (1971). Experiences with a sales districting model: criteria and implementation. *Management science*, P41-P54.

Kim, C. E., & Anderson, T. A. (1984). Digital disks and a digital compactness measure. *Annual ACM Symposium on Theory of Computing, Proceedings of the sixteenth annual ACM symposium on Theory of computing* pp. 117-124). New York, N.Y.: ACM Press.

Li, H. (1989). Spatio-temporal pattern analysis of managed forest landscapes: a simulation approach.

Li, W., Church, R. L., & Goodchild, M. F. (2014). The *p*-compact-regions problem. *Geographical Analysis*. (forth-coming)

Li, W., Goodchild, M. F., & Church, R. L. (2013a). An efficient measure of compactness for 2D shapes and its application in regionalization problems. *International Journal of Geographical Information Science*.

Li, W., Li, L., Goodchild, M. F., & Anselin, L. (2013b). A Geospatial Cyberinfrastructure for Urban Economic Analysis and Spatial Decision-Making. *ISPRS International Journal of Geo-Information, 2*, 413-431.

Longley, P., Goodchild, M. F., Maguire, D., & Rhind, D. (2005). *Geographic information systems and science*: Wiley.

Massam, B. H., & Goodchild, M. F. (1971). Temporal trends in the spatial organization of a service agency.

Canadian Geographer, 15, 193-206.

McDonald, J. F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics, 21*, 242-258.

McGarigal, K., & Marks, B. J. (1995). Spatial pattern analysis program for quantifying landscape structure. *Gen. Tech. Rep. PNW-GTR-351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station*.

Miller, H. J., & Shaw, S.-L. (2001). *Geographic information systems for transportation: Principles and applications*: Oxford University Press.

Minor, S. D., & Jacobs, T. L. (1994). Optimal land allocation for solid-and hazardous-waste landfill siting. *Journal of Environmental Engineering, 120*, 1095-1108.

Osserman, R. (1978). Isoperimetric Inequality. *Bulletin of the American Mathematical Society, 84*, 1182-1238.
Osserman, R. (1978). Isoperimetric Inequality. *Bulletin of the American Mathematical Society, 84*, 1182-1238.
Ricca, F., & Simeone, B. (2008). Local search algorithms for political districting. *European Journal of Operational Research, 189*, 1409-1426.

Ríos-Mercado, R., & Salazar-Acosta, J. (2011). A GRASP with strategic oscillation for a commercial territory design problem with a routing budget constraint. *Advances in Soft Computing*, 307-318.

Ríos-Mercado, R. Z., & Fernández, E. (2009). A reactive GRASP for a commercial territory design problem with multiple balancing requirements. *Computers & Operations Research, 36*, 755-776.

Sorensen, P., & Church, R. (2010). Integrating expected coverage and local reliability for emergency medical services location problems. *Socio-Economic Planning Sciences, 44*, 8-18.

Tung, C. P., & Chou, C. A. (2007). Application of tabu search to ground water parameter zonation. *JAWRA Journal of the American Water Resources Association, 38*, 1115-1125.

Williams, J., & Re Velle, C. (1996). A 0-1 programming approach to delineating protected reserves. *Environment and Planning B, 23*, 607-624.