Assessing Continuous Demand Representation in Coverage Modeling

Ran Wei[1], Alan T. Murray[1]

[1]GeoDa Center for Geospatial Analysis and Computation, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287-5302, USA
Email: {ranwei; atmurray}@asu.edu

1. Introduction

Minimizing costs and maximizing coverage are common goals in many planning contexts. Examples include locating fire stations to guarantee immediate response to calls for service, placing emergency warning sirens to alert the public of impending danger, or siting cellular towers to allow widespread access of wireless broadband. To support these planning problems, spatial optimization problem are utilized. This necessitates an abstraction of both the planning problem as well as geographic space. Unfortunately, such abstraction results in unintended errors when traditional spatial optimization coverage models are applied. In order to reduce coverage errors caused by spatial abstraction, a number of geographic representation schemes have been proposed. One intriguing approach relies on vector GIS based overlay as a way to identify the finest level of geographic resolution needed for a demand region in order to eliminate representation errors. However, this involves many GIS/geometric operations, such as polygon overlay and partitioning that are well-known to be computationally intensive (De Berg et al. 2008). Combined with issues of potential facility locations, it is possible that the resulting number demand units using overlay is excessive and beyond computational capabilities. This paper therefore investigates the operational and computational challenges of polygon overlay for representing continuous demand in coverage models, an issue that has yet to be explicitly studied. The analysis results provide insight regarding expected problem sizes and computation requirements if this is relied upon in coverage modeling.

2. Coverage model

An important spatial optimization problem is the location set covering problem (LSCP) (Murray and Wei 2013). The intent of the LSCP is to site the minimum number of facilities in order to ensure complete coverage of a demand region. It was first formulated by Toregas et al. (1971) to site emergency service facilities. Consider the following notation:

i = index of demand units to be served (entire set I, $|\mathrm{I}| = \mathrm{M}$);

j = index of potential facility locations (entire set J, $|\mathrm{J}| = \mathrm{N}$);

$$a_{ij} = \begin{cases} 1, & \text{if demand i can be suitably served by facility j;} \\ 0, & \text{otherwise;} \end{cases}$$

$$\Omega_i = \{ j | a_{ij} = 1 \};$$

$$X_j = \begin{cases} 1, & \text{if a facility is sited at potential location j;} \\ 0, & \text{otherwise;} \end{cases}$$

The a_{ij} elements indicate an evaluated coverage standard, reflecting suitable serve response in distance or travel time. Taking fire services as an example, $a_{ij} = 1$ would reflect personnel at a fire station at location j being able to reach demand at location i in 8 minutes or less. GIS is generally used to evaluate such service standards (Church and Murray 2009). Given this notation, the LSCP formulation follows:

Location Set Covering Problem (LSCP)

$$\text{Minimize} \quad Z = \sum_{j} X_j \tag{1}$$

$$\text{Subject to} \quad \sum_{j \in \Omega_i} X_j \geq 1, \forall i \tag{2}$$

$$X_j = \{0,1\}, \forall j \tag{3}$$

The objective of the LSCP, (1), is to minimize the number of facilities located. Constraints (2) ensure that each demand is covered by at least one sited facilities. Constraints (3) impose binary integer restrictions on decision variables. It should be noted that the potential facility locations, demand units and coverage sets need to be identified in advance to apply the LSCP. Doing so in a manner that is error free remains a challenge (Murray and Wei 2013).

3. Spatial representation and coverage overlay

When the purpose of planning is to provide complete service to a continuous region, there is generally a need to abstract the continuous region into discrete spatial objects, like points or polygons. However, the abstraction process is well known to create uncertainties or errors in coverage modeling. As an example, a point representation could result in an underestimate of actual required facilities, while an area representation may lead to an overestimate of needed facilities (Murray and Wei 2013). Recently, a vector based overlay approach was employed to partition the continuous demand region, where each resulting unit is a disjoint piece of coverage overlay (Cromley et al. 2012). Figure 1 shows how this approach works when the facility coverage is measured by Euclidean distance. Given that each demand unit is the smallest areal unit that a sited facility could possibly cover, there is no partial coverage of demand. That is, a demand unit is either completely covered or not covered at all. Such a property is extremely important, as this ensures that demand unit polygons are partitioned in a way that no error in coverage would result. Thus, the optimal solution for the corresponding LSCP instance will reflect the true minimum number of facilities required to cover the entire region. This is theoretically important. However, an issue is then whether the approach is computationally feasible in practice, as the complexity of geometric computations involved are not trivial nor would the resulting spatial optimization model necessarily be possible to solve.

4. Evaluating coverage overlay

There are two major computational concerns with overlay approach. The first is attributed to the a priori generation of demand units, involving considerable geometric operations and processing. The second is associated with solving the resulting LSCP. It is well recognized that the number of polygons arising from overlay is a good estimate for overlay processing time (Saalfeld 1989). The input for overly in this case involves the coverage region for each potential facility location, *J*, and the demand region. The size of LSCP model largely determines whether it can be solved using commercial software. The size of an LSCP is dictated by the number of generated demand units to be covered (derived in this case using overlay) as well as the number of potential facility locations. The number of generated demand units, therefore, is a proxy for assessing the computational performance of the coverage overlay approach.

In general, it is difficult to predict the number of polygons that result from an overlay operation (NCGIA 1997). Actually, even establishing a valid bound is not easy (Saalfeld 1989). However, if the coverage standard is the distance r, and facility coverage is considered to be a circle with radius r, then an upper bound for the number of overlay polygons does exist. This assumption is not unrealistic because many types of facilities have circular service coverage, like emergency warning sirens and cellular towers. Equivalent to the plane division by circles problem in Yaglom and Yaglom (1987), the maximum number of demand units into which facility coverage can be divided is:

$$\mathrm{M} \leq \mathrm{N}^2 - \mathrm{N} + 2 \tag{4}$$

where M is the number of demands and N is the number of potential facility locations, consistent with previous terminology. While the above bound may be useful in assessing the number of demand units that would be generated, it could be a very loose bound in practice. As an example, 291 facilities, shown in Figure 1b, could possibly generate 84,392 demand units theoretically. In practice, however, only 13,320 units are observed. The bound is 7 times larger than the actual number. As a result, there is a need to derive a more accurate estimate for the number of demand units. Given that more intersections of coverage circles will typically result in more generated demand units, it is reasonable to account for the potential intersections in evaluating generated demand units. To do so, the pairwise distances between potential locations must be computed in advance, which is computationally efficient as facility sites are represented as points. Let $\mathrm{d_{jj'}}$ be the distance between facility locations j and $\mathrm{j'}$, r the coverage radius, and Ψ_j the set of locations whose coverage intersects with the coverage of facility j, the average number of intersections per coverage (ANI), can be defined as follows:

$$\Psi_\mathrm{j} = \left\{\mathrm{j'} \middle| \mathrm{d_{jj'}} \leq 2\mathrm{r}\right\} \tag{5}$$

$$\mathrm{ANI} = \frac{\sum_\mathrm{j} |\Psi_\mathrm{j}|}{\mathrm{N}} \tag{6}$$

The ANI value is an average, independent of the number of potential facility locations. If both variables are combined to estimate the number of demand units, a better prediction is expected.

5. Preliminary results

In order to evaluate the computational challenges posed by the overlay based approach, 54 application instances are examined. The details of study design will be included in the full paper. A standard regression model is established including the two explanatory variables that are hypothesized to be related to number of demands generated (M). Specifically, this involves the number of potential locations (N) and the average number of intersections per coverage (ANI) as follows:

$$\sqrt{M} = \beta_0 + \beta_1 * \mathrm{N} + \beta_2 * \mathrm{ANI} \tag{7}$$

The regression results are presented in Table 1. Both of the explanatory variables are significant with p-value≈0.00. The coefficients are positive, implying that larger number of potential locations and average intersections will result in more demand units. This conforms to the initial hypothesis. The R-square is close to 99%, which is encouraging as it suggests about 98.92% of the variability in predicting new observations can be explained using this model, along with the approximately 99.1% of the variability in the original data. The predictive capability of the model is satisfactory and it is reliable to use this model to predict the coverage overlay partition results.

Table 1: Regression results for number of demands using number of sites and ANI

Variable	Coefficient	Std. of coefficient	t-statistics	p-value
Constant	215.56	2.359	91.38	0.000
Number of sites	60.624	3.132	19.36	0.000
ANI	134.477	3.132	42.93	0.000
F-statistics	2852.2 (p-value = 0.000)			
R-square	99.1%			
R-square (prediction)	98.92%			

6. Future work and conclusion

The paper has established theoretical and practical bounds for the number of generated demand units that can be expected from polygon overlay. This is important in spatial optimization approaches that rely on overlay for pre-processing. The analysis demonstrates that it is possible to accurately predict the number of demand units that will be generated by coverage overlay using two easily accessible variables. However, this is only one part of the evaluation. More work remains for the processing and solution time to formalize the computational complexity involved in coverage overlay approach. The findings of this paper can be used to determine whether this approach is computationally feasible for a given practical application, contributing to addressing abstraction and representation issues in coverage modelling.

References

Cromley, R.G., Lin, J. and Merwin, D.A., 2012, Evaluating representation and scale error in the maximal covering location problem using GIS and intelligent areal interpolation. *International Journal of Geographical Information Science*, 26:3, pp. 495-517.

De Berg, M., Cheong, O., Van Kreveld, M. and Overmars, M., 2008, *Computational geometry: algorithms and applications* (Berlin, Heidelberg: Springer).

Church, R.L. and Murray, A.T., 2009, *Business Site Selection, Location Analysis and GIS* (New York, NY: Wiley).

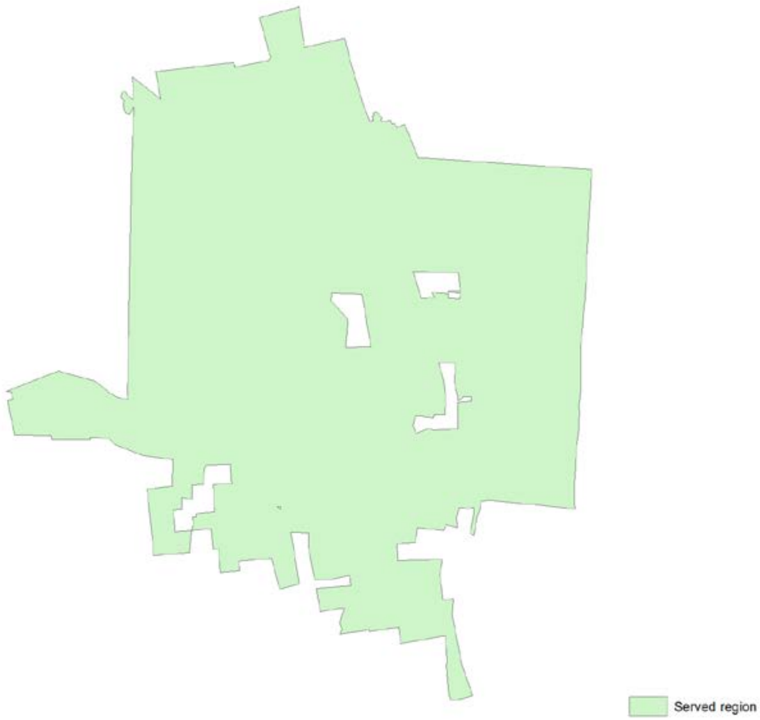
Murray, A.T. and Wei, R., 2013, A computational approach for eliminating error in the solution of the location set covering problem. *European Journal of Operational Research*, 224:1, pp. 52-64.

NCGIA, 1997, The polygon overlay operation. Available online at: http://www.ncgia.ucsb.edu/giscc/units/u186/u186.html.

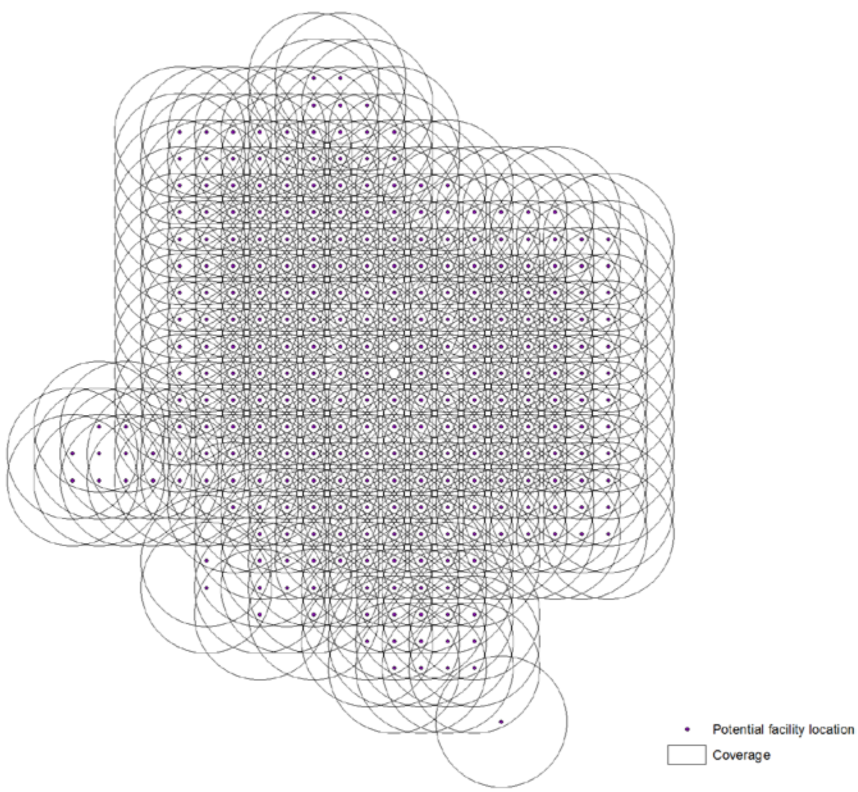
Saalfeld, A., 1989, The Combinatorial Complexity of Polygon Overlay. In *Ninth International Symposium on Computer-Assisted Cartography*, ASPRS (Ed.) (Baltimore, MD: The Society), pp. 278-288.

Toregas, C., Swain, R., ReVelle, C. and Bergman, L., 1971, The location of emergency service facilities. *Operations Research*, 19, pp. 1363-1373.

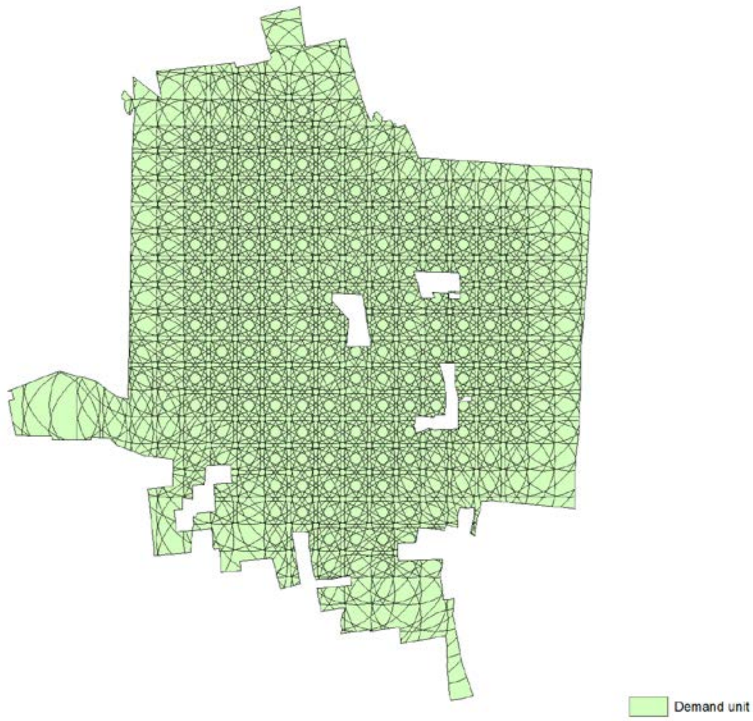
Yaglom, A.M. and Yaglom, I.M., 1987, *Challenging mathematical problems with elementary solutions* (Mineola, NY: Dover Publications).



(a)



(b)



(c)

Figure 1: Coverage overlay approach. (a) Served region; (b) Potential facility locations and their corresponding coverage; (c) Demand units created by coverage overlay.