Geocomputation and Spatial Analytics

A. T. Murray*[1]

[1]Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106, USA
*Email: amurray@ucsb.edu

Abstract

Geocomputation and spatial analytics have a shared history, and often times the distinction between the two areas is less than clear. While both occupy a place under the banner of GIScience, the emergence of spatial data science as a more broadly consuming grouping perhaps makes any distinctions less significant. Irrespective of the overarching label or naming preference, there are lessons to be learned about specification, representation, implementation and interpretation, all of which have implications for openness. This paper provides a comparative overview, delving into the nuances of geocomputation and spatial analytics. This is done to make a number of points associated with recent trends in open spatial data science.

Keywords: GIS, modelling, spatial optimization, open geospatial science

1. Introduction

The evolution of geographic information system (GIS) capabilities, data capture mechanisms and readily available / accessible digital information has been remarkable. As a result, the needs for various sorts of analysis has grown incredibly. This no doubt helps to explain the emergence of specialty areas like geocomputation and spatial analytics. Geocomputation is often defined as computational methods capable of accounting for special characteristics of spatial data through the use of artificial intelligence (and machine learning) approaches, such as genetic algorithms, neural networks, cellular automata, agent-based modelling, etc. A pioneering force was Stan Openshaw, and the modelling context of geocomputation extended to most forms of geographical analysis, like optimization, statistics and clustering (Gahegan 1999, Harris et al. 2017, Thill and Dragicevic 2018). Dominating much of geocomputation is what can be referred to as heuristic techniques, where simulation and other ad hoc procedures are employed to mimic and replicate processes as well as project the future and prescribe remedies. The work of Clarke et al. (1997) and Waddell (2002), and others along these lines, come to mind as very representative of developments in this area, where artificial intelligence (or machine learning, simulation, etc.) oriented approaches rely on a learning mechanisms to establish rules by which change occurs. The rules are then analysed in various ways and/or used to derive a solution of some sort.

One of my first Geocomputation conference contributions was in 1999 (Virginia, USA) involving the use of one such artificial intelligence approach, cellular automata, subsequently published as Ward et al. (2000). My involvement in other geocomputational efforts has continued over the years, including Ward et al. (2003), Murray et al. (2008), Kim et al. (2008), Tong et al. (2009), Wei and Murray (2014) and Hong et al. (2017). Drawing on Murray (2010, 2017), Murray (2019) suggests that spatial analytics include all quantitative methods that support analysis, policy and planning involving geographic space. My view is that geocomputation has historically been more narrowly defined in terms of associated methods, but the semantics are of secondary importance here. What I would note is that, as suggested in Thill and Dragicevic (2018), GIS has made significant headway in providing access to spatial analytics and geocomputation. This becomes increasing relevant in a number of ways. In what follows I review a number of insights associated with specification, representation, implementation and interpretation, all relevant in the context of openness.

2. Specification

Method specification can be challenging in many ways. Since my primary area of specialization is spatial optimization, a field falling under both spatial analytics and geocomputation, I will make a number of preliminary points before more broad generalisations. A discussion of spatial optimization can be found in Tong and Murray (2012). Consider the following definition:

Spatial optimization - structuring and solving a problem involving decisions to be made that must conform to various restrictions imposed, where decisions, coefficients, relationships and constraints may be spatial and/or aspatial in nature.

This is a reasonably broad, but vague, description. Contrast this with the following:

Minimize $f(X)$ (1)

Subject to $g_i(X) \geq b_i \quad \forall i$ (2)

where X is a vector of decision variables, x_j, $f(\quad)$ a function relating inputs, $g_i(\quad)$ is the i^{th} function relating inputs, and b_i a coefficient associated with each constraint i. The objective of this optimization model, (1), seeks to minimize the function of decision variables. The constraints, (2), impose restrictions on decision variables through functions $g_i(\quad)$ in relation to its right hand side value b_i. The spatial nature of the problem would arise through some combination of decision variables (e.g., X), coefficients, functions that are geographically specific (e.g., $f(\quad)$ and $g_i(\quad)$), and/or spatial structure that is imposed by constraints, (2). Further, we will assume that some approach exists for solving this model.

The point to be made is that spatial optimization becomes more explicit through the formulation, (1)-(2), in contrast to the written description given above. Increased specificity beyond this is certainly possible as well. Consider the following:

Minimize $\sum_j c_j x_j$ (3)

Subject to $\sum_j a_{ij} x_j \geq b_i \quad \forall i$ (4)

$x_j = \{0,1\} \quad \forall j$ (5)

where c_j is a coefficient associated with each decision variable in (3), the objective, a_{ij} is a coefficient associated with each decision variable j and i^{th} constraint (4), and all other notation as previously defined. Binary conditions on decision variables have been amended to the model in (5). This formulation of a spatial optimization problem is now much more specific, though for generalization c_j, a_{ij} and b_i have been used to denote coefficients rather than actual observed numerical values. However, in order to solve this model these a priori defined and known values would be necessary to specify the application. Further, under certain conditions it is likely that optimization software, commercial or open source, may be able to able to solve an associated problem instance given the linear functions involved.

From a specification perspective, an issue with spatial analytical or geocomputational approaches is whether the problem/model has been rigorously and/or sufficiently formalized. A mathematical model is essentially the gold standard because of the specificity and communication involved. The assumptions, relationships, etc. are explicit in the mathematical statements used to express the model/problem. An observation regarding geocomputational methods is that most can be characterized as vague in some manner, lacking explicit mathematical specification. For example, Clarke et al. (1997) summarized their cellular automaton model using the following pseudo code:

```
Read data layers
Initialize random numbers & control parameters
For n iterations {
        For t time periods {
        Apply change rules
        Apply self modification rules
        Compute and save descriptive statistics
        }
}
Write images
```

Waddell (2002) used a table to compare operational model characteristics of the developed simulation approach, UrbanSim, to other urban models. In addition, a figure to depict data integration as well as a figure to diagram model structure and processing are provided. This more or less represents the “model”. Both Clarke et al. (1997) and Waddell (2002) typify a number of geocomputation approaches in that there is considerable ambiguity in many aspects of what the model is doing. Questions that come to mind are the many. What are the planning or management units? What are the options for change and/or modification (the decisions)? What spatial relationships are considered? What are the transition probability functions and/or rules? Many other questions could be added to this list.

Interestingly, vague details are also offered for many spatial analytical heuristics as well as methods provided in GIS packages, such as ArcGIS and TransCAD. Murray (2018a) and Murray et al. (2019) report that a host of spatial optimization models are available in these packages, yet descriptions available in the software or through online help are not mathematical, employ differing/inconsistent naming conventions, and generally do not reference any associated literature.

The important message here is that a lack of specificity for any quantitative method is highly problematic in a number of ways. A fundamental issue is that vagueness leads to uncertainty. When this happens, replication, verification, openness, etc. is not possible. Beyond this, it is unclear exactly what a user is doing, which therefore limits capabilities to defend what is being done. So, non-explicit approaches are not necessary wrong or incorrect, but rather create uncertainties that could impact the utility of a method / model.

3. Representation

A particularly acute issue in spatial analytics and geocomputation is the significance of the geographic representation of a region under study. Openshaw and Taylor (1981) refer to the modifiable areal unit problem (MAUP), suggesting that the results of a particular quantitative method may be the byproduct of spatial unit definition and/or geographic scale of analysis. That is, depending on the method used, the finding obtained may be biased or impacted by the representation of a geographic area. An interesting take on this is offered in Tobler (1989), indicating that the issue may well be attributable to the method utilized, not geographic representation per se. Specifically, if the method is sensitive to geographic representation, then it is frame dependent. The implication is that a frame dependent method, according to Tobler (1989), is the wrong method in terms of structure and/or conceptualization. Tobler (1989) suggested that researchers seek out and develop frame independent methods, where an analytical approach would produce consistent results irrespective of spatial representation.

Regardless of whether you prefer the MAUP perspective or frame dependence characterization, what is clear is that a range of spatial analytic and geocomputational approaches are indeed sensitive to spatial representation. There has in fact been much research highlighting this issue as well as development of methods that are less susceptible or free from MAUP issues. Murray (2018b) discusses work in spatial optimization, where a coverage location problem is considered. It was demonstrated that through the use of GIS, theoretical properties could be established for eliminating MAUP, with a strategy devised to guarantee an optimal solution. That is, Murray (2018b) shows that the considered coverage problem can be structured and solved using a frame independent approach.

Representation is an important concern for spatial analytic and geocomputational methods. One issue that comes to mind is how can representation impacts be assessed in a more formalized manner? To date, approaches for evaluating MAUP, if any, have been ad hoc in spatial analytics and geocomputation. While most quantitative methods are believed to be impacted by MAUP, evaluation and assessment in a given application context is necessary. A second issue is, assuming that MAUP or frame dependence is detected, how to address representation impacts? The advice now is simply to use the finest spatial resolution possible, reflecting that this is essentially the best that can be done in many circumstances. Of course, this is not particularly scientific nor does it address the more overarching and pressing issues of methodological rigor associated with spatial representation.

4. Implementation

The implementation or solution of models in spatial analytics and geocomputation is often unclear, ambiguous or vague. Spatial optimization, as an example, is highly dependent on model specification, formally or informally. The above mathematical models are a formal means of specification. Of course, spatial optimization is also about model or problem implementation, or solution. The entire point of model specification is that it is a precursor to solution. Solving a spatial optimization may be accomplished in two ways, exactly or heuristically. Murray (2010) and Tong and Murray (2012), among others, provide discussion on problem solution. An exact method is one capable of identifying an optimal solution, where it is guaranteed to be the best and can be proven as such. There is no better solution that can be identified using any other imaginable approach. On the other hand, a heuristic method is one comprised of rules or ad hoc procedures for solving a problem. The identified solution may be good, feasible, useful, etc., but there is no guarantee that it is optimal. Heuristics are often relied upon because they are computationally efficient, are easy to understand and/or implement, find good solutions, offer fast solution times, are not dependent on proprietary software, etc. There are indeed appealing reasons for employing a heuristic to solve a problem. Nevertheless, a heuristic cannot guarantee that an optimal solution is found.

In terms of implementation (or solution), spatial analytics and geocomputational approaches are regularly solved using software and libraries lacking clarity in what has been done or the methods used. That is, it is not always evident whether the solution approach is exact or heuristic. Exact methods are preferable for solving any model because one can rely on associated optimality properties. Of course, this may not always be possible as the model may not be amenable to exact solution, it may be too difficult to solve, solution may require an excessive amount of time, proprietary software is too expensive, etc. These are points made in Murray (2018b) and Murray et al. (2019). The implementation issue clearly centers on what has been done, whether an exact or heuristic approach has been used, the parameters utilized, assumed conditions, etc. The reason is that all of these issues may impact solution quality but also have implications for significance. For example, if heuristic results are obtained, it may be prudent to establish confidence bounds on findings in some manner. Murray et al. (2019) note that a final solution for a spatial optimization model in GIS, as an example, is simply passed along to the user with no mention of certainty, yet often the solution is often found using a heuristic technique. Additionally, many of these heuristics do not allow for parameter adjustment or control, further hindering evaluation of solution quality.

5. Interpretation

Cognitive skills are something that varies considerably among individuals. It should not therefore be surprising that individual ability to interpret spatial analytic and geocomputation results would vary as well. Clearly this is challenging give the complexities of specification, representation and implementation already noted, but there is even more to it than this. Irrespective of whether one implements an exact or heuristic approach in code or relies on GIS or other software, the fact remains that user skill levels vary. People may or may not have formal training in a particular area, such as spatial statistics, clustering, spatial optimization, etc. This will no doubt impact their ability to interpret findings. This does not necessary mean they will or will not utilize a method or set of methods. What is true is that point-and-click access makes it easy for even the most novice of users. Indeed, this is by design. Point-and-click is a software system that is user friendly where the computer mouse / touch pad can be used to point to a function and click on it (or through the use of menu pull downs) thereby launching the method, likely providing an easy to navigate interface of the model / method. Most GIS packages are accessible in this manner, but also a range of software that supports spatial analytics. Again, the ability to point-and-click does not necessarily translate to being capable of interpreting the results in a valid, meaningful and significant manner. Finally, gaining problem insights is often essential in spatial analytics. The goal is to seek out implications, where this can then enable planning, management and policy decision making. This is the essence of interpretation.

Spatial analytics and geocomputational approaches face significant challenges in the area of facilitating interpretation. As noted above, a major difficulty is differing technical skills and training in advanced spatial analytical methods. Beyond this, there are a range of issues worthy of serious consideration. For example, how does one interpret model findings / results in the context of parameters and assumptions, geographic and model based abstraction, data quality and uncertainty, etc. Additionally, there are indeed issues associated with specification and implementation, such as interpreting significance for non-optimal results. Does it providing meaningful insights? Can the results be used to draw a valid inference? Are the implications generalizable? There are no doubt others as well. An interesting issue involving geographic context is that spatial analytics and geocomputation often produce results involving spatial configuration. How does one interpret spatial configuration when models have alterative optima? That is, many spatial optimization models are known to have multiple alternative optima, meaning that spatial configuration can be very different between alternative solutions yet all would be optimal (equivalent objective value). Of course, if a heuristic is used then there may be close to optimal spatial configurations as well. Finally, in the case of a multi-objective model, there are possibly many non-dominated solutions (Pareto optima), all of which represent optimal solutions for given user preferences associated with trade-offs. Interpretation in each of the above cases is difficult at best.

6. Discussion and Conclusions

The intent of the paper has been to raise specification, representation, implementation and interpretation issues in spatial analytics and geocomputation. An obvious challenge remains for improved communication. However, it is clear that further research is necessary on ways to compare, assess, summarize, visualize, etc. in the context of various types of uncertainty. To be sure, uncertainty arises due to specification, representation, implementation and/or interpretation facets encountered. Figure 1 depicts a solution to a spatial optimization problem, the LSCP, obtained using ArcGIS ("Minimize_Facilities" option). This particular application is detailed in Murray et al. (2019), and involves siting California Redemption Value centres. The solution identified in Figure 1 indicates that 42 service centres are required. However, given the coverage standard of 0.5 miles, it is actually possible to serve all indicated demand (supermarkets) with only 41 centres. The spatial configuration is different, but more importantly 42 centres is more costly than 41. There are fixed costs as well as recurring annual costs associated with personnel, transportation, etc. Non-optimality due to the use of a heuristic (implementation) in this case is actually a major issue, likely to translate to millions of dollars in unnecessary expenditures. How can users, agencies, decision making bodies exercise fiduciary responsibility without better knowledge?

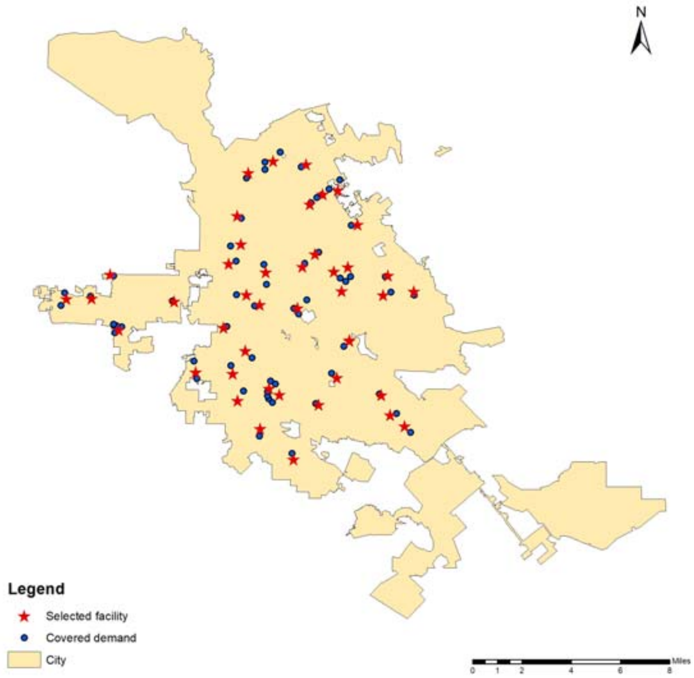

Figure 1: Heuristic solution for a location coverage problem.

With recent trends moving toward open geospatial science, reproducibility and replicability, the results reflected in Figure 1 raise significant concerns. Because of uncertainty associated with methods (specification), the results do not reflect openness. As the solution is identified using a heuristic process, it may in fact not be reproducible nor replicable, per se.

Open geospatial science no doubt has an important role in the future of geocomputation (and spatial analytics). It reflects notions that all are welcome to contribute along with transparency, replicability and reproducibility, among others, in areas where spatio-temporal detail is inherently meaningful (Sui 2014, Rey 2018). Open geospatial science, or perhaps open spatial data science, is the culmination of decades of efforts to make programming, analysis, modelling, planning, etc. better in a variety of ways. One aspect of this is not reinventing the wheel for each and every project. Data, code, models, algorithms, programs, software, etc. should have some general utility beyond its original specific purpose. Reuse is a good rationale but there is also elimination of error, where reproduction always has the potential to introduce bugs, assuming something was done correctly to begin with. Either way we are talking about a savings of time and effort in carrying out implementation or conducting analysis. Nevertheless, open science also has to do with communication as well as expectations of replicability and reproducibility.

Specification, representation, implementation and interpretation concerns have been raised associated with spatial analytical and geocomputation methods. Perhaps the biggest contrast between spatial analytics and geocompuation is the tendency to rely more on heuristic approaches in geocomputation, but there are clearly many instances of where this is also true in spatial analytics. Of more significance is that improved capabilities to communicate quality and uncertainty is more critical than ever. Research is needed along the lines outlined given recent trends and the growing significance of spatial data science.

7. References

Clarke, K.C., Hoppen, S. and Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and Planning B: Planning and design, 24(2), pp.247-261.

Gahegan, M., 1999. Guest editorial: what is geocomputation? Transactions in GIS, 3(3), 203-206.

Harris, R., O'Sullivan, D., Gahegan, M., Charlton, M., Comber, L., Longley, P., Brunsdon, C., Malleson, N., Heppenstall, A., Singleton, A. and Arribas-Bel, D., 2017. More bark than bytes? Reflections on 21+ years of geocomputation. Environment and Planning B: Urban Analytics and City Science, 44(4), pp.598-617.

Hong, I., Kuby, M. and Murray, A., 2017. A deviation flow refueling location model for continuous space: A commercial drone delivery system for urban areas. In Advances in Geocomputation (eds D. Griffith, Y. Chun, D. Dean), pp. 125-132. Springer.

Kim, K., Murray, A.T. and Xiao, N., 2008. A multiobjective evolutionary algorithm for surveillance sensor placement. Environment and Planning B: Planning and Design, 35(5), pp.935-948.

Murray, A.T., 2010. Quantitative geography. Journal of Regional Science, 50(1), pp.143-163.

Murray, A.T., 2017. Regional analytics. The Annals of Regional Science, 59(1), pp.1-13.

Murray, A.T., 2018a. Complexities in spatial center derivation. Transactions in GIS, 22(6), pp.1335-1350.

Murray, A.T., 2018b. Evolving location analytics for service coverage modeling. Geographical Analysis, 50(3), pp.207-222.

Murray, A.T., 2019. Sustainability and resilience through micro-scale decisions for change. Draft manuscript.

Murray, A.T., Matisziw, T.C., Wei, H. and Tong, D., 2008. A geocomputational heuristic for coverage maximization in service facility siting. Transactions in GIS, 12(6), pp.757-773.

Murray, A.T., Xu, J., Wang, Z. and Church, R.L., 2019. Commercial GIS location analytics: capabilities and performance. International Journal of Geographical Information Science, 33(5), pp.1106-1130.

Openshaw, S. and Taylor, P.J., 1981. The modifiable areal unit problem. In 'Quantitative Geography: a British View' (eds N. Wrigley, R. Bennett), pp. 60–69. Routledge and Kegan Paul, London.

Rey, S.J., 2018. Code as text: Open source lessons for geospatial research and education. In GeoComputational Analysis and Modeling of Regional Systems (pp. 7-21). Springer.

Sui, D., 2014. Opportunities and impediments for open GIS. Transactions in GIS, 18(1), pp.1-24.

Thill, J.C. and Dragicevic, S. (eds.), 2018. GeoComputational Analysis and Modeling of Regional Systems. New York: Springer.

Tobler, W.R., 1989. Frame Independent Spatial Analysis. In 'The Accuracy of Spatial Databases' (eds M. Goodchild, S. Gopal), pp. 115-122. New York: Taylor and Francis.

Tong, D. and Murray, A.T., 2012. Spatial optimization in geography. Annals of the Association of American Geographers, 102(6), pp.1290-1309.

Tong, D., Murray, A. and Xiao, N., 2009. Heuristics in spatial analysis: a genetic algorithm for coverage maximization. Annals of the Association of American Geographers, 99(4), pp.698-711.

Waddell, P., 2002. UrbanSim: Modeling urban development for land use, transportation, and environmental planning. Journal of the American planning association, 68(3), pp.297-314.

Ward, D.P., Murray, A.T. and Phinn, S.R., 2000. A stochastically constrained cellular model of urban growth. Computers, Environment and Urban Systems, 24(6), pp.539-558.

Ward, D.P., Murray, A.T. and Phinn, S.R., 2003. Integrating spatial optimization and cellular automata for evaluating urban change. The Annals of Regional Science, 37(1), pp.131-148.

Wei, R. and Murray, A.T., 2014. A multi-objective evolutionary algorithm for facility dispersion under conditions of spatial uncertainty. Journal of the Operational Research Society, 65(7), pp.1133-1142.