Chapter 10
GIS in Regional Research

Alan T. Murray

10.1   Introduction

Geographic information systems (GIS) have come to be an important component of regional science. This is not particularly surprising given the very characteristics used by Walter Isard in establishing and defining regional science (see Isard 2003). Namely, the use of rigorous analytical methods stemming from multiple disciplines coming together to study real world problems and issues was generally noted as foundational principles of regional science. Such approaches no doubt must be supported by data of some sort, likely with a spatial/geographic orientation. And it turns out that spatial data are messy and complicated, requiring specialized techniques, methods, processes, etc. devoted to the creation and use of such data, but also that spatial data often contain a wealth of implicit knowledge.

With the origins of regional science in the 1950s, there was a rather simplistic view of geographic space and the phenomena associated with it. Computers were just coming onto the scene and were far from accessible. Those fortunate enough to get access encountered major computing limitations due to hard drive, memory and processor components. A simplified view of a region or city was a necessity, with objects of interest often being a point with one or more attributes. Further, the number of points was traditionally kept to a minimum, again because computing capabilities were limited. A common analytical processing need was to derive distance, and often Euclidean (or straight line) travel was deemed sufficiently representative of spatial interaction between two points.

Fast forward to present day. Computing is an afterthought, if even considered at all. Computing power and supporting software exists on laptop and handheld

A.T. Murray
Department of Geography, University of California at Santa Barbara, Santa Barbara, CA, 93106, USA
e-mail: amurray@ucsb.edu

devices (and watches). Data are plentiful, often with more data generated in real time than can possibly be synthesized and understood. Monitoring devices are everywhere: woven into clothing; on bracelets and cellular phones; embedded in vehicles; scanning sensors recording purchases and spending behavior; and satellites circling the earth measuring and recording any and everything. Regions and cities are not assumed to be points but rather more complex areal features that can be highly non-homogeneous in terms of attribute characteristics, response and behavior. Limitations on the number of observations considered are substantially relaxed as it is not uncommon to consider thousands or millions of spatial objects, often managed and processed using GIS. And many sorts of geographic interactions may be present, including agglomeration, unique paths of travel between objects, neighboring impacts, etc.

Interests in regional science have evolved accordingly as well. A unit of analysis is no longer assumed to be homogenous and static, but rather varied and changing over time. We have witnessed a systematic reduction of simplifying assumptions and a recognition of the importance of local detail. It is no surprise then that the power and capabilities of GIS have only gained significance. In many ways, GIS is becoming, or has become, a central approach used in regional science. However, there remains a view that GIS is merely a tool for making maps. This chapter sets out to provide an overview of GIS. In doing this, the intent is to highlight current and future capabilities beyond map making, as GIS is specifically designed to deal with geographic data creation and the analysis of this data.

10.2   GIS

A formal definition of GIS is that it is a combination of hardware, software and procedures that support spatial analysis and decision making. GIS necessarily requires capabilities for data capture, management, manipulation, analysis and display associated with spatially referenced data (see Church and Murray 2009; Longley et al. 2015). Collectively, this means that geographic space can be abstracted as layers of information, as suggested in Fig. 10.1, enabling integration and analysis within and across layers.

The process of data capture in GIS involves abstracting the earth, or a portion of it, as a digital representation. This is often done as either a raster or a vector model. The creation of data is possible using many approaches, possibly involving the use of GPS, aerial sensing, drones and/or other sensing devices or may be based on manual digitizing, automated conversion and/or geocoding. The data management component of GIS is concerned with storage, access and query efficiency. The operational response and processing capabilities of GIS software is dependent on managing data efficiency.

The manipulation of data in GIS is necessary for many reasons. Transformation of different layers of information to a consistent frame, or coordinate system, is a very common manipulation approach. Another classic spatial manipulation
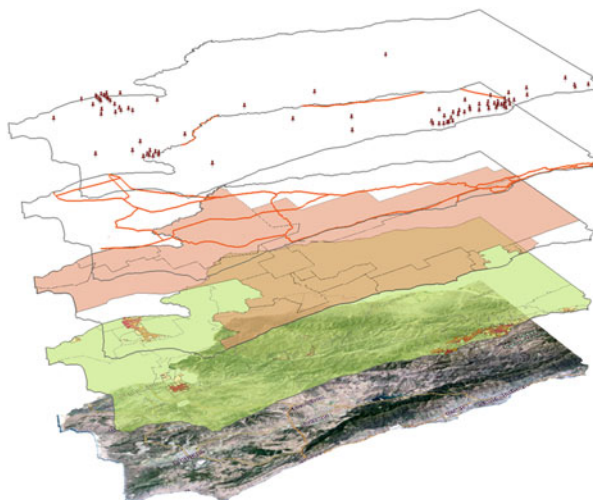
Fig. 10.1 GIS based layers of spatial information

approach is projection of three-dimensional latitude and longitude referenced spatial information into a two-dimensional coordinate system. Processing, calculation and display, as an example, may require two-dimensional representation, depending on the analysis setting. Other contexts, however, may require a three-dimensional depiction. Various spatial manipulation approaches are considered standard in GIS, including simplification and aggregation, among others. Examples of each of these manipulation functions in regional science work is readily found. Simplification may involve the derivation and use of a centroid to represent a county. Aggregation might entail the spatial combination of two (or more) adjacent census tracts in order to form one new polygon that represents the unit of analysis.

The data analysis capabilities of GIS have historically been perceived to be limited. However, this is a bit unfair because there are actually many analytical capabilities, ranging from attribute summary, spatial summary, containment assessment, polygon overlay, map algebra, deriving distance and proximity, buffering, interpolation, cluster detection, etc. In total, there is a wealth of analytical capabilities, but has historically not included advanced statistical, geostatistical, geosimulation and spatial optimization approaches (Anselin and Getis 1992; Fischer and Nijkamp 1992; Goodchild and Haining 2004). The major commercial packages generally include some access to geostatistical and spatial optimization methods, and libraries/software like GeoDA offer advanced spatial statistical methods (Church and Murray 2009; Murray 2010).

Finally, the display of spatial data in GIS has been a mainstay. Often this has entailed the making of a map, either on screen or in a paper form. More contemporary approaches have emerged to support geovisualization in 2D and 3D, but also account for some aspects of temporal variability (see Maciejewski 2014; Rey 2014; Longley et al. 2015).

10.3   Representation and Data

GIS is special and important because the data managed corresponds to activities and observations that exist/occur on the surface of the earth, and also because it has the capability to interact, query, manipulate, etc., associated information in various ways. Accordingly, representation issues are critical as the earth is not regularly shaped nor easily specified in a digital environment. While a convenient, simplified assumption is to consider the earth to be a sphere or an ellipsoid, it actually is neither. Often a geoid may do a reasonable job of approximating the earth. Nevertheless, there are always challenges in accurately and appropriately representing the surface of the earth. Depending on the scale, location of interest and the purpose of a study, a particular representation and associated datum(s) of the earth may be reasonable. Care must therefore be taken with the underlying representation of the earth to ensure that it is satisfactory for intended usages. Associated with an assumed representation of the earth is the need for a referencing system so that attributes and characteristics of places may be encoded, processed and analyzed. Referencing may depend on the abstraction of geographic space, such as whether the interest is in objects or fields. Objects typically consist of points, lines and polygons, whereas a field is generally a regular discretization of continuous space, such as a raster grid surface. The implication of referencing and objects/fields are many, but historically two factors have figured prominently: storage and processing efficiency. System response and computing needs are dependent on how data are stored. Access and query of information are intrinsically linked to the storage of data.

Vast amounts of geographic data exist in a range of formats. Spatial information and sources for obtaining it can be found in Church and Murray (2009) and Longley et al. (2015), among others. Various attempts have been made or exist that bring together publicly available spatial information, and are referred to as Geolibraries or Geoportals (Longley et al. 2015). Some are the byproduct of federal, state and/or local government efforts to ensure public access. An example at the federal level is DATA.GOV. At the state level, California provides public access to geospatial information through the state geoportal, http://gis.ca.gov/, as an example. At a local level, agencies like SANDAG (http://www.sandag.org/) in San Diego provide varying levels of access to certain geospatial data. Other communities, cities and states have policies and data access portals meant for public consumption of geographic information.

Historically the US Census has supplied important data about people and the economy in the United States. To do this, the Census employs an army of people, with primary products being the Decennial Census of Population and Housing (every 10 years), Economic Census (every 5 years), Census of Governments (every 5 years) and the American Community Survey (annually). Of course, a valuable component of census data is that digital records are available for at least a recent history. While a very good source of information, there are issues with the data. These issues can impact data quality, reliability, spatial and temporal accuracy, etc. Particular issues include sampling bias, undercounts, variable ambiguity,

conflation, reporting delay/change, as well as others. From a spatial perspective, the fact that census unit boundaries can change presents significant challenges, and most importantly introduces further data uncertainty. Resolving attribute values for reporting units across time periods means that various types of interpolation (intelligent estimation/guessing) are necessary.

A wealth of spatial data now is obtained from sensing based platforms. This includes aerial and ground based equipment ranging from Global Positioning System (GPS), satellites, aircraft and drones to stationary and mobile video, images, road counters and other sensors. While GPS, satellite imagery and aircraft LiDAR are particularly commonplace and accessible, emerging technological capabilities provided by drones offer potential for real time and continuously updated remotely sensed information. On the ground, sensing equipment and technology abounds, from Google Street View vehicles to red light cameras to security video to activity detection devices, there is arguably more continuous sensor data than can be processed and ingested.

Of course, one source of spatial information is to obtain it from private data vendors, typically a byproduct of an assimilation effort on the part of the vendor where various data are brought together through the scraping of digital and print resources. Vendors such as Nokia (HERE), Walls and Associates (National Establishment Time Series), Nielsen (PRIZM), etc. turn raw data into valuable spatial information, often associated with the location of public and private goods or services. Worth noting in particular is a significant reliance on geocoding in the creation of vendor data. An example is National Establishment Time Series produced by Walls & Associates that effectively converts Dun and Bradstreet establishment data into digital, spatially referenced information. This is done by interpreting the establishment/company street address as a global position. This is known as geocoding, the formal process associated with taking a local street address reference and identifying geographic coordinates for that address on the surface of the earth, namely, a latitude and longitude (Murray et al. 2011). While a very common process to produce digital information, there are a range of issues associated with such data. Geocoding works by identifying a successful address match in a street centerline database. Often match rates are high with most commercial software, but not perfect. You can expect 5–10% of the address data to not be successfully matched. Beyond this, a successful match does not necessarily translate to spatially accurate information. The reason for this is that address matching involves interpolation along street centerline segments to estimate the location of an address number. Further, an offset distance is assumed to put the point on the building, hopefully a "rooftop hit". Ultimately, little is often known about the actual spatial accuracy of geocoded data as the located point may not be precisely on the house, business or building, nor necessarily in the associated land parcel, neighborhood block or census tract. Errors in positional accuracy of a few meters to a few kilometers are not unusual. Worth mentioning is that business address data can be problematic. Often, records reflect headquarters only as a registered place of business, but information on where employees undertake the work is not known.

Another class of spatial information is generated by individuals, possibly solicited or unsolicited. This includes what is widely known as volunteered geographic information (VGI). Websites and software that facilitate VGI include WikiMapia, OpenStreetMap and Map Maker, where individuals create, collect and disseminate spatial data (Longley et al. 2015). Of course, other sources of VGI could include Twitter feeds (when location is disclosed or inferred), Yelp, Urbanspoon, etc. Noteworthy regarding such data is that it may be biased in many ways, not reflective of all opinions, not representative of all social classes, lacking consistency and objectivity, and may not have extensive spatial coverage. Further, data standards and associated metadata often are lacking in many ways. Other sources of user generated data are rather indirect sources, perhaps unknowingly provided by an individual. Spatial location, time and behavior can be obtained through the use of cellular phones and other electronic equipment as well as through the use of customer loyalty card programs, among others. Cell phones are typically GPS enabled, or location can be inferred from cellular towers and satellites. Customer loyalty card programs represent a growing source of data where companies like dunnhumby, Aimia, emnos, Nielsen, Symphony EYC, 5one and Demandtec employ analytics to better understand our collective behavior and trends. While not necessarily publicly available at this time, the data and information extracted by cellular providers and companies with loyalty cards can be purchased and used in various ways without any need for consent on the part of individuals.

10.4 Significance of GIS

There are many implications for regional science in the growth and evolution of GIS. As suggested above, geographic data availability across a range of domains changes how processes may be considered and the detail at which it can be conceived. Not only with respect to more traditional concerns associated with residential location, as an example, but now detailed information on movement and mobility patterns throughout a day or week. Beyond this, there is ubiquitous monitoring by sensors on the ground and in the sky. Big data associated with objects and group or individual activities is generated daily, if not hourly. In many respects this enables details about place and behavior to be considered, and also accounted for explicitly in analysis and modeling efforts. While GIS does have considerable mapping and analytical capabilities, the use of GIS based data and methods to support advanced mathematical and statistical modeling continues to be noteworthy. Thus, what makes GIS special is the ability to create, manage and use data in order to derive and exploit spatial knowledge. Given that the data managed by GIS is geographic in nature, there are many spatial relationships that result. Of note are proximity, adjacency, connectivity, shape, direction, containment, concentration and scale. From a database perspective, GIS is interesting and unique because these spatial relationships are often implicit in that they are not computed in advance and

stored as part of the database. Rather, these relationships may be inferred because of geographic location and derived on the fly as needed.

There are many examples of regional science based work that has made use of, or integrated with, GIS in order to carry out advanced modeling of some sort. A prominent area of work is associated with land use planning. Supporting this is the use and development of cellular automata approaches; and more recently, agent-based methods. These approaches operate using some representation of a region, often a regular raster, combined with current and past data on development in each cell. Working on a rather informal model specification, rules are established that reflect observed or inferred growth patterns that can be used to mimic or estimate future patterns. Work in this area includes that of Clarke et al. (1997) and Ward et al. (2000), among many others. What is noteworthy is that spatial relationships have been found to be key to developing good land use transition/change rules. For example, land use around a given area is particularly influential, but also current and future infrastructure is an indicator of likely land use change. The advancement of cellular automata and other related approaches for land use planning/analysis has gone hand-in-hand with GIS proliferation and access to more detailed data. Approaches have pursued greater specification and linkage to regional models (Ward et al. 2003) but now also account for more features of land cover dynamics. Recent discussion of these approaches can be found in Clarke (2014).

A prominent approach in regional science is assessment, evaluation and/or detection of activity concentration. They may represent clusters, hot spots, cold spots, neighborhoods, homogeneous response zones, etc. Identification of such areas could be associated with response correlation or simply detecting agglomeration of some sort. Factors associated with housing prices or foreclosure could be of interest, as an example. Alternatively, one may want to detect whether there are high rates of activity, such as crimes or industry mixes. To support this, a variety of methods have been developed and applied in regional science, including local and global measures of spatial autocorrelation as well as scan statistics. Spatial information critical to most approaches is neighborhood structure, the so called weights matrix or a scanning window depending on the methodological inquiry. This is often based on adjacency or proximity relationships. GIS is invaluable in deriving spatial relationships and details along these lines. A recent review of work in this area can be found in Murray et al. (2014), but of note is the use of GIS to specify more spatially relevant relationship structures, such as AMOEBA and LOSH (Getis 2015) and irregular scan "windows" (Murray et al. 2014).

The direct measurement of proximity is fundamentally important to most regional science based investigations. The interest is generally associated with the spatial interaction between two locations, and in particular the shortest path. This may represent a travel route taken by an individual or a cost/distance to travel. When travel is not restricted to a road network, this is a continuous space problem that offers an infinite number of travel path options. A popular choice for quantifying proximity is to assume straight line (Euclidean) travel between two locations. This can be problematic for many reasons but often does not reflect an actual travel path. Greater realism means that obstacles must be taken into account, such as water
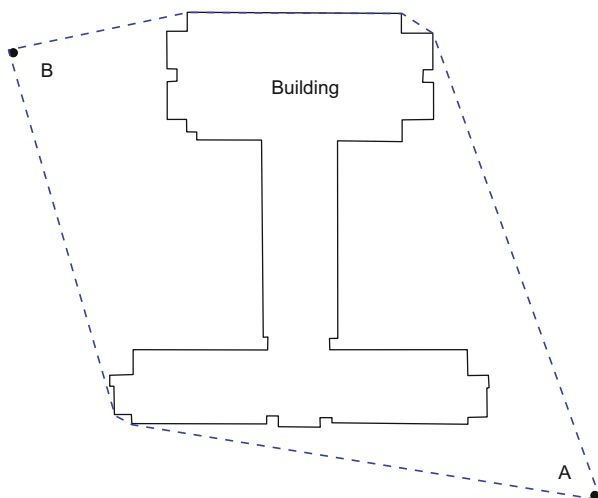
Fig. 10.2 Convex hull for associated objects (A, B and Building)

bodies, bridges, structures, airports, mountains, canyons, etc. When travel is not permitted through obstacles and/or buildings, the most efficient continuous space route is known as the Euclidean shortest path. Of course, travel through geographic space necessarily requires spatial information about movement options between a given origin and destination. Beyond this, however, standard operations in GIS dealing with spatial proximity and computational geometry are critical. In fact, Hong and Murray (2013) report an approach to identify an optimal Euclidean short path based on the use of a convex hull, readily identified using GIS. Figure 10.2 depicts the convex hull associated with three objects, two points and a building. The convex hull represents a minimum length boundary containing all three objects. What Hong and Murray (2013) proved is that the shortest path lies on this boundary, assuming travel from point A to point B, effectively reducing the infinite number of routes possible through continuous space to only a finite number of polygon segments along the convex hull. A technique based on convex hulls was generalized for the case of multiple obstacles. The significance of this is that one can identify an obstacle avoiding shortest path in real-time using GIS, enabling navigation and wayfinding as well as providing an ability to more accurately model travel behavior.

A fairly common spatial analytical method used in regional science is a location model. In particular, coverage models have proven to be invaluable for addressing many types of service situations (Murray 2016). Whether the circumstances involve prescriptive plans or a descriptive understanding of an existing system, location coverage models enable mathematical specification and solution derivation when facilities are to be placed throughout a geographic service area. Facilities could correspond to fire stations, clinics, cell towers, restaurants, etc. The geographic service area may be based on travel time, distance, visibility and/or audibility, and may be regular or irregular in shape. GIS is, therefore, invaluable for helping to
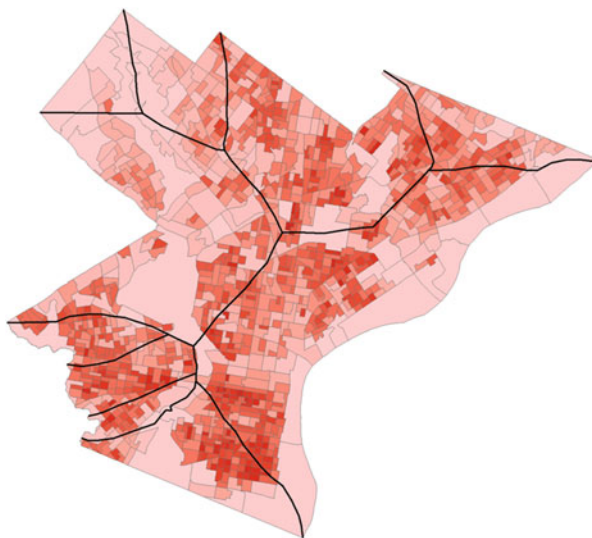
Fig. 10.3   Skeleton of a polygon region

structure proximity, adjacency, contiguity, concentration, etc. in associated models. Beyond this, Murray and Tong (2007), Murray et al. (2008) and Matisziw and Murray (2009) have demonstrated that spatial knowledge and relationships can be exploited through the use of GIS. In particular, Murray and Tong (2007) derived finite dominating sets corresponding to locations where an optimal configuration of facilities would be limited to for continuous space coverage problems. GIS facilitates the identification by systematic evaluation of service areas using overlay functions, reducing an infinite number of siting possibilities to a finite set. Murray et al. (2008) and Matisziw and Murray (2009) proved that an optimal facility site would be located along the skeleton (or medial axis) of a region. As an example, a region is shown in Fig. 10.3 with demand for service distributed throughout. The skeleton for this region is also shown, and we know that service coverage is maximized when the facility is sited somewhere along the skeleton. Again, GIS enables this property of a region with respect to coverage to be exploited in various ways, providing a means of solutions based on the derived property.

10.5   Conclusions

This chapter serves many purposes. One is to provide an overview of GIS within the context of regional science. Beyond this, the hope is that a characterization of GIS based features will result in a greater understanding of what has been and could be done to support regional science using GIS. Finally, a number of recent developments associated with the integrated use of GIS in regional science were discussed. Speculative discussion based on this overview are now possible.

On the analysis side, much of what was discussed in this chapter reflects the continued convergence of GIS and spatial analytics approaches, something noted in Goodchild and Haining (2004), among others. This can be seen in much of contemporary regional science work in that GIS is central to all facets of a study, beginning with data management all the way to the application of particular analytic approaches. The future of regional science will be one where GIS is increasingly more central. This no doubt presents many research challenges for addressing issues of integration as well as efficiently deriving/solving associated models.

What is abundantly clear at this time is that uncertainty abounds. GIS highlights that this is, in fact, the case with digital information. Positional location of objects stored in GIS is rarely without error or uncertainty. Further, even in cases where there is a high level of positional certainty, various manipulation operations could create resident uncertainty. A similar observation holds for model abstractions that attempt to mimic observed regional systems and behavior. The model is a simplification of an actual system, omitting certain features and nuances. Add to this the fact that we may have a limited or biased understanding of systems, processes and behaviors, then collectively there is much potential for all sorts of direct and indirect uncertainty. As a result, this will continue to force researchers to rethink and reevaluate how we approach regional science, and more importantly how we can address issues of uncertainty and bias in the many forms that it may arise. While GIS may highlight how resident uncertainty exists in data, it by no means offers a roadmap on how to take uncertainty into account nor how analysis and planning can be bolstered.

Given technological and computing advances, it is clear that big data will continue to change what is done in regional science and what can be done. The level of detail at which data are collected will necessitate changes in applied analytics, often reflecting a relaxation of simplifying assumptions that have long been relied upon. There is little doubt that this is a good thing as it will enable better modeling and analysis to be carried out. This will translate into better insights, improved plans and superior policy development. The challenges, of course, are how methods will evolve accordingly. This will change perspectives and understanding. As a result, fundamental assumptions likely will prove problematic, thereby needing to be relaxed and/or modified.

While somewhat related to the previous point(s), it is a fact that changes over time are really not well understood in general. Aspects of land use planning may be an exception, as noted above. However, there are significant challenges for relating change over time to actual behavior/response/operation. This can be said for past and current systems, but also for future conditions. What will residential land use patterns look like? To what degree will employment centers continue to decentralize? How will travel and behavior patterns change? What technological advances can we expect and what are implications for regional systems? These questions and others simply highlight that cross sectional work has real limitations, yet with a greater ability to use more detailed spatial information will come an enhanced capability to simulate future conditions.

In closing, this chapter has demonstrated the significance of GIS in regional science to date, but likely this is merely a starting point. What GIS really tells us at this point is that assumptions regarding sufficient data quality are actually very problematic. The impacts and implications for analysis and planning are actually not well understood at all. Couple with this increasing amounts of detailed data from a variety of sources over time, and the suggestion is that there is much left to do in regional science.

References

Anselin L, Getis A (1992) Spatial statistical analysis and geographic information systems. Ann Reg Sci 26:19–33

Church RL, Murray AT (2009) Business site selection, location analysis and GIS. Wiley, New York

Clarke KC (2014) Why simulate cities? GeoJournal 72:129–136

Clarke KC, Hoppen S, Gaydos L (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environ Plann B Plann Des 24:247–261

Fischer MM, Nijkamp P (1992) Geographic information systems and spatial analysis. Ann Reg Sci 26:3–17

Getis A (2015) Analytically derived neighborhoods in a rapidly growing West African city: the case of Accra, Ghana. Habitat Int 45:126–134

Goodchild MF, Haining RP (2004) GIS and spatial data analysis: converging perspectives. Pap Reg Sci 83:363–385

Hong I, Murray AT (2013) Efficient measurement of continuous space shortest distance around barriers. Int J Geogr Inf Sci 27:2302–2318

Isard W (2003) History of regional science and the Regional Science Association International. Springer, Berlin

Longley PA, Goodchild MF, Maquire DJ, Rhind DW (2015) Geographic information systems and science, 4th edn. Wiley, New York

Maciejewski R (2014) Geovisualization. In: Fischer MM, Nijkamp P (eds) Handbook of regional science. Springer, Berlin, pp 1137–1155

Matisziw TC, Murray AT (2009) Siting a facility in continuous space to maximize coverage of a region. Socio Econ Plan Sci 43:131–139

Murray AT (2010) Quantitative geography. J Reg Sci 50:143–163

Murray AT (2016) Maximal coverage location problem impacts, significance, and evolution. Int Reg Sci Rev 39:5–27

Murray AT, Grubesic TH, Wei R (2014) Spatially significant cluster detection. Spat Stat 10:103–116

Murray AT, Grubesic TH, Wei R, Mack EA (2011) A hybrid geocoding methodology for spatio-temporal data. Trans GIS 15:795–809

Murray AT, Matisziw TC, Wei H, Tong D (2008) A geocomputational heuristic for coverage maximization in service facility siting. Trans GIS 12:757–773

Murray AT, Tong D (2007) Coverage optimization in continuous space facility siting. Int J Geogr Inf Sci 21:757–776

Rey SJ (2014) Spatial dynamics and space-time data analysis. In: Fischer MM, Nijkamp P (eds) Handbook of regional science. Springer, Berlin, pp 1365–1383

Ward DP, Murray AT, Phinn SR (2000) A stochastically constrained cellular model of urban growth. Comput Environ Urban Syst 24:539–558

Ward DP, Murray AT, Phinn SR (2003) Integrating spatial optimization and cellular automata for evaluating urban change. Ann Reg Sci 37:131–148

Alan T. Murray is professor, Department of Geography, University of California at Santa Barbara (UCSB). His primary research interests are geographic information science; spatial optimization; health informatics; urban growth and development; land use planning; urban, regional, and natural resource planning and development; and, infrastructure and transportation systems. He previously held academic appointments at Drexel University, Arizona State University and Ohio State University. Previous faculty positions were at Drexel University, Arizona State University and Ohio State University. Dr. Murray earned a Ph.D. in geography from the University of California at Santa Barbara in 1995.