Evolving Spatial Analytics and Rural Area Classifications

Alan T. Murray

Center for Spatial Analytics and Geocomputation

College of Computing and Informatics

Drexel University

Philadelphia, PA 19104, USA

Email: amurray@drexel.edu

Abstract

This chapter contributes to an ongoing discussion about rural area classification. Reviewed in this chapter are spatial analytic methods for carrying out analysis, planning, decision making or policy formulation involving rural areas in some manner. Of particular focus is how geographic data and spatial analytic methods have involved in various ways. There have been considerable and sustained advances in computing and processing capabilities, making big data and real time analytics the norm for studies and evaluation. This means that it is now possible to work with spatially disaggregate data in ways that accommodate varying interpretation of rural classification, shifting from a one size fits all approach pursed in the past. Of course, challenges remain as spatial analytic methods are recognized as potentially being biased or otherwise influenced by how geographic data is represented as well as data quality issues. More pressing concerns in rural classification are likely spatial data quality and error issues in analytical contexts.

Introduction

There is much interest in rural areas in the United States, and elsewhere. The motivation for this varies substantially. Ultimately, however, the reasons for the interest have to do with inherently challenged conditions of poverty, health, housing, education, employment and the like that characterize some rural communities. To this end, there is a critical need for understanding these conditions and the factors that have contributed to their existence, with the hope that solutions can be found that bring about positive change, particularly in extreme cases or where rural conditions differ significantly from urban areas.

While conceptually there is an understanding of what rural means, there are many different formal specifications of the defining characteristics of a rural area. This is not really surprising because how one would define rural would be dependent on the interests, concerns and purpose of who is examining and/or questioning existing or hypothesized conditions. Examples of differing definitions of what it means to be rural can be inferred from US federal agencies, such as the US Census Bureau, the White House (Office of Management and Budget), the US Department of Human and Health Service (Office of Rural Health Policy) and the US Department of the Treasury (Internal Revenue Service), among others. This chapter will not delve into this issue, but rather simply note that there is no single accepted definition nor necessarily a definition that would satisfy all circumstances. What can be said is that rural areas are characterized by low population densities and are away from urban areas. The Economic Research Service (ERS) (US Department of Agriculture) categorizes areas along a rural-urban continuum at the county level (http://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation.aspx) using an ordinal scale from 1 to 9. On one end of this continuum , a county is “completely rural or less than 2,500 urban population, not adjacent to a metro area” (code 9). On the other end of the continuum, a county is urban, or code 1, for “counties in metro areas of 1 million population or more”. An important distinction from the rural definition of the US Census Bureau is that a Block is the spatial unit being categorized (in contrast to the much larger county), and is either rural or urban making it binary in nature (not a continuum). However, ERS does offer rural classification delineations at finer resolutions, such as the FAR (half kilometer grid aggregated to ZIP codes).

Why differences in rural definitions? Isserman (2005) notes that the varying concerns articulated by different state/federal agencies with respect to rural areas, like the US Economic Development Administration, the Appalachian Regional Commission, the US Department of Agriculture (Rural Community Empowerment Program), the US Department of Housing and Urban Development, and the US Census Bureau, play an important role in how rurality is viewed. There is no doubt this is true. As an example, the US Department of Treasury (Internal Revenue Service) defines rural with respect to airports based upon commercial airline passenger departures of less 100,000 per year (http://www.irs.gov/publications/p510/ch04.html) in the context of the Essential Air Service program (see Grubesic et al. 2013 for more details). Interestingly, this definition is not explicitly based on population density at all. Given the broad interests in rural conditions, it is not surprising that there are many definitions of what it means to be rural.

This chapter is conceived with the above context in mind, that defining rural in a formal, quantitative manner is challenging. However, advances in computing, geographic data and spatial analytics point to the need for continued evaluation of formal specification of rural. The next section reviews geographic data, and recent data generation trends. This is followed by a formal presentation of a number of spatial analytic techniques, including GIS (geographic information systems), location based measures, metrics and multipliers, spatial statistics, spatial optimization, geosimulation and ESDA/SDSS. The modifiable areal unit problem (MAUP) and frame dependence are then discussed, highlighting that spatial analytics are known to be sensitive to the underlying representation of geographic space. The implication being that both data issues and methods applied matter in various ways, so technical issues can impact analytical findings relied upon to identify harmful rural conditions, allocate help, aid and resources, development plans for change, and formulate rural policy. The chapter ends with discussion and concluding comments.

Geographic Data

A wide array of geographic data exists and is relied upon to carry out analysis of all sorts centered on or related to rural areas. General discussion of 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 a variety of 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, the Commonwealth of Pennsylvania provides public access to geospatial information through PASDA (http://www.pasda.psu.edu/), which is part of National Spatial Data Infrastructure (http://www.fgdc.gov/nsdi/nsdi.html) sponsored by US Department of the Interior. At a local level, the City of Philadelphia provides access to certain geospatial data through its Office of Innovation and Technology as a part of its Open Data initiative (http://www.phila.gov/it/Innovation/OpenData/Pages/GeospatialData.aspx, see also https://www.opendataphilly.org/). Other communities, cities and states have similar policies and data access portals.

Historically the US Census has served to supply 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 and the American Community Survey. 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 and do 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 a significant challenges, and most importantly introduce further data uncertainty. Resolving attribute values for reporting units across time periods means that various types of interpolation are necessary, and by definition an interpolation method is a guess or estimate.

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. Of course, ERS already makes use of some data of this type, relying on impervious surface composition in the rural-urban continuum code assignment, as an example. However, there is much more than this that exists. While GPS, satellite imagery and aircraft LiDAR are particularly commonplace and accessible, emerging technological capabilities provided by drones offers 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, a prominent source of spatial information is available from private data vendors, typically involving the assimilation of various data sources or scraping digital and print sources of data. 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 into good spatial accuracy. 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, and may be worse for rural areas. In fact, Cayo and Talbot (2003) found positional error for rural addresses to be nearly seven times that of urban areas. Worth mentioning as well, business address data may be complicated to begin with, possibly representing only headquarters and not regional offices, reflecting a registered place of business but employees undertaking the work elsewhere, and other quirks that may be industry specific.

Another class of spatial information is individual user generated. This includes what is widely known as volunteered geographic information (VGI). Web sites and software that facilitate VGI include WikiMapia, OpenStreetMap and Map Maker, where individuals create, collect and disseminate spatial data (Goodchild 2007). Of course, other sources of VGI could include Twitter feeds (when location is disclosed or inferred), Yelp, Urbanspoon, etc. Noteworthy points 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 many not have extensive spatial coverage. Further, data standards and associated metadata often is 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.

Given the above sources of spatial information, there is little doubt we have entered the age of big data, where volume, velocity and variety of spatial data is arguably beyond current capabilities to analyze or even store in some cases. What is exceedingly clear is that there are significant differences in data availability for rural areas. Some places lack cellular coverage or are hampered by topological and environmental conditions that result in service gaps and degradation, interrupted GPS signals, and the like. Of course, behavioral patterns are different from people in urban areas, resulting in technological adoption differences as well varying economic practices. In particular, trade, bartering and other aspects of the informal economy are all too common in rural areas. By design they are not amenable to contemporary digital tracking, possibly skewing observed behavior through other sources of data.

Spatial Analytics

The broad collection of spatial analytics have come to include those quantitative methods utilized to support analysis, policy and planning involving geographic information, where a range of approaches can be relied upon for carrying out systematic inquiry. Such a definition is consistent with quantitative geography and geocomputation detailed in Murray (2010a) and Openshaw (2014). Often spatial analytics would include geographic data creation methods like remote sensing, but for our purposes here the characterization will be limited to: geographic information systems (GIS); measures, metrics and multipliers; spatial statistics; spatial optimization; geosimulation; and exploratory spatial data analysis (ESDA) and spatial decision support systems (SDSS). Each will now be described in detail.

Geographic Information Systems (GIS) - Formal descriptions of GIS note that it is the combination of hardware, software and various procedures to support spatial analysis and decision making. This involves data capture, management, manipulation, analysis and display that is unique to spatially referenced data (see Church and Murray 2009, Longley et al. 2015).

Data capture in GIS involves abstracting the real world as a digital representation, often as either a raster (regular or irregular tessellation of space) or a vector (objects consisting primarily of point, lines and/or polygons) model. Data creation may involve the use of GPS, aerial sensing, drones or other ground based sensing devices, as noted previously, or may involve digitizing, conversion, geocoding, etc. Given the above discussion on geographic data, we will not delve further into aspects of data capture, but rather note that there are a range of approaches that are utilized for generating spatial information.

Data management in GIS is primarily concerned with storage, access and query efficiency. The response and capabilities of a commercial GIS software package is clearly dependent on data management efficiency across each of these concerns. While enhanced computing capabilities have contributed to increased GIS processing power, advancements in various geographic data management components have served to make big data contexts possible and a rather typical operational context for contemporary GIS.

Data manipulation in GIS is possible in a variety of ways. It is very common to project 3-dimensional (3D) latitude and longitude referenced spatial information into a 2-dimensional (2D) coordinate system, constituting a classic spatial manipulation operation. Other spatially oriented manipulation approaches in GIS include simplification, aggregation, disaggregation and interpolation, among others. Examples of simplification include using a centroid to represent a county, and the use of a street centerline to reflect a multi-lane street, freeway or interstate. Aggregation is fairly common in GIS. An example of aggregation is combining counties to form a metropolitan area. A mainstay of GIS has been the capability to deal with different spatial unit geographies as well as manipulate them in various ways depending upon a study context.

Data analysis in GIS has historically been viewed as limited, but it most certainly is much more than mapping. Among the many analysis capabilities associated with GIS are attribute summary, spatial summary, containment assessment, polygon overlay (vector), map algebra (raster), deriving distance and proximity, buffering, interpolation, cluster detection, etc. In the end then, there are in fact may different analytical capabilities that are standard in GIS, and they support spatial investigation, exploration and assessment.

Finally, data display in GIS centers on map making, both on screen and as paper oriented products. What has emerged over the past decade is advanced capabilities for geovisualization in 2D and 3D, and in some cases over the temporal dimension as well. This includes internet/web and user centered approaches (see Peng and Tsou 2003; Shyy et al. 2014).

The implications for rural classification in light of GIS (and a variety of spatial information) is that processing software and computing capabilities enable individual level analysis, and can be done in near real time. Accommodating changes and variations in rural definition is therefore relatively manageable and technically possible. Beyond this, an important point is that digital geographic information is imprecise in various ways. It begins in abstraction processes where the real world is approximated in a digital environment and continues through GIS processes and functions facilitating data creation, manipulation and analysis. These points will be re-visited later in the paper.

Measures, Metrics and Multipliers – Particularly common in the study of rural and urban conditions is the reliance and application of a range of measures, metrics and multipliers. These quantitative summary approaches are structured to derive insights about issues or phenomena of interest.

Popular in the social sciences are the dissimilarity index and Gini coefficient, but also measures of segregation. Some measures or indices are simple counts combined in some way, while others may account for relationships with other variables or spatial features. Examples include the location quotient, the human development index, accessibility/remoteness index, isolation, segregation, exposure, etc. Recent reviews can be found in Grubesic and Murray (2008), Wong and Shaw (2011) and Osth et al. (2015). Effectively what can be noted is that there are indices or measures that attempt to account for almost any nuance of interest. Specific to rural research, the index of rurality is another such example, where the measure is based on principal components analysis to integrate the combination of different factors. Waldorf (2006) structures the index of relative rurality to be continuous, scaled and flexible through the integration of population density, urban extent and remoteness. A recent review specific to rural work can be found in Caschili et al. (2015).

The rural-urban continuum codes too can be considered an index or measure in the sense that a county is evaluated based on metropolitan status, population, proximity as well as impervious surface and then a code is assigned. The code is an ordinal measure.

Summary is even possible in a more complicated mathematical system. This is the premise of matrix techniques like eigenvalues and eigenvectors. They reduce multi-dimensional information to a single number or vector, respectively. Meaning is then attached to this resulting summary measure. In regional analysis it is very common to utilize input-output analysis, computable general equilibrium models, etc. to elicit summary measures, or rather multipliers as the case may be. Details on select methods can be found in Grubesic and Murray (2004) and Fischer and Getis (2009), among others.

There no doubt are implications for rural classification and the analysis of rural issues. The first thing is that indices and measures continue to evolve in different ways. This is a function of better data, enhanced geographic detail, and greater insight. On the other hand, in many ways an issue remains that such approaches lack statistical significance, may have weak or no theoretical justification and may be sensitive to data quality.

Spatial Statistics – There are many statistically based methods that have been relied upon to support rural inquiry. This includes classic statistics like correlation, analysis of variance, regression, etc. However, geographic data has proven to be special in various ways, but most notably due to the fact that observed attributes/conditions are not independent but rather are similar in some way. Such a relationship is precisely the so called first law of geography attributed to Waldo Tobler, that all things are related but nearby things are more related than those further way. As a result, approaches for spatial sampling and spatial statistics (and in some sense geostatistics) have emerged as specialized statistical methods for appropriately dealing with geographic space/information. Examples include approaches capable of evaluating point patterns, measure spatial correlation, account for dependency in regression model, etc. Specific models and/or methods are nearest neighbor, quadrat, kernel density, k functions, clustering, the expansion method, spatial autoregressive models, geographically weighted regression, etc. Reviews of select methods can be found in Rogerson (2009), Anselin et al. (2013) and Murray et al. (2014), among others.

There are implications for rural classification and rural study. Such methods tend to be established within a context of inference, enabling a statement of statistical significance. Of course, this is necessarily predicated on assumptions of some sort, and they may or may not be appropriate or restrictive in some situations. On the positive side, these methods continue to evolve and be refined in various ways. An example of this is spatial autocorrelation. The initial thrust in measuring and evaluating spatial autocorrelation was establishing an appropriate measure. Some of the more prominent regional (or global) approaches are Moran's I, Geary's c and G. Subsequent work in the 1990's moved toward developing local measures of spatial autocorrelation, examples of which include Gi* and the general decomposition of global measures explored in Anselin (1995) referred to as LISA statistics. More recently is the recognition that how we specify proximity relationships through the use of a weights matrix, W, is very important, and methods like AMOEBA are necessary to help detect/justify such relationships (Getis 2015).

Spatial Optimization – Much analytic work on understanding conditions in rural areas is predicated on evaluating proximity and access to important goods and services. Often this has involved comparison to standards as well as examining system efficiency. Carrying this out has meant that optimization methods are needed to assess how well observed spatial patterns or services compare with a theoretical best case. Beyond understanding and evaluation, spatial optimization approaches are regularly applied to allocate resources and site services in order to maximize access and accessibility as well as minimize costs. This has involved the development and application of linear, integer and dynamic programming models that reflect a problem or situation of interest, possibly solved by exact or heuristic methods. Examples in geographic contexts include the use of spatial interaction approaches, spatial analytics and location models, network analysis and path derivation. Reviews of methods, models and approaches associated with spatial optimization can be found in Church and Murray (2009), Murray (2010b) and Yao and Murray (2014).

A spatial optimization model has a range of spatial components, and could include decisions on where something should be located, service allocation, route specification, district design, etc. Similarly, constraining conditions might include geographic needs and/or restrictions, perhaps requiring separation, maximum spacing, rapid response, etc. The coefficients of a model may be geographic or proximity based in nature. Attributes of desired service/design outcomes may be shape or pattern oriented, perhaps elongated, fragmented or compact, depending on needs or desires. Finally, spatial relationships may be central, such as adjacency, connectivity, contiguity, containment, intersection, nearest, furthest, etc. Discussion of these characteristics and other aspects of spatial optimization may be found in Tong and Murray (2012).

The implications for rural classification and analysis are likely similar to those already noted. As is the case for spatial statistics, there is the possibility to attribute significance to findings. Specifically, if exact methods are used then it is possible to establish efficiency or quality. Alternatively, there are assumptions that are implicit, and findings may be sensitive to data quality.

Geosimulation – The area of simulation in a geographical context has been referred to in many different ways. Here the term geosimulation is adopted, and is meant to include geocomputation and spatial microsimulation. This area of analytics formalizes processes in various ways, then seeks to mimic stochastic and random elements of evolution and behavior in order to come up with scenarios or outcomes associated with a particular issue/concern. A common area of application is examining urban growth and development as well as land use change. Important elements associated with this are behavior, movement and interaction patterns. Issues addressed have included gentrification, sprawl and rural/urban migration. Particular techniques utilized to simulate processes include cellular automata, agent based models and neural networks. Reviews of work in this area can be found in Ward et al. (2000) and Huang et al. (2014), among others.

The implications for rural classification are that identified scenarios may or may not likely. Often there is a lack of statistical significance. Further, model(s) and processes may not be well justified in terms of an underlying theory. Similarly, geosimulation approaches may be less than rigorously defined, relying on a substantial degree of parameter specification. Such specification is often subjective in nature.

ESDA and SDSS – Exploratory spatial data analysis (ESDA) and spatial decision support systems (SDSS) represent an organized and concerted effort to systematically investigate geographic information, problem context, insights and assumptions using a combination of analytical methods, primarily those noted above. Often this is geographic (or map) based interaction and display where GIS is playing a central role. ESDA focuses on knowledge discovery and hypothesis specification/formalization, and may or may not include confirmatory methods. In contrast, SDSS is typically structured to support decision making of some sort, in an exploratory fashion. Discussion of ESDA and SDSS can be found in Murray (2010a,b), Anselin (2012) and Murray et al. (2012).

Some interesting implications for rural classification stemming from ESDA and SDSS are that there are dependencies on user/analyst interaction to insert insights in the pathway towards knowledge discovery. This, of course, is both good and bad.

Application Issues

The previous section has detailed a range of spatial analytic methods that have been utilized and applied to address a variety of rural (and urban) issues. While some implications for application have been noted, a number of overarching issues have not yet been discussed. To this end, this section discusses in particular the modifiable areal unit problem (MAUP) and frame dependence. Both are a recognition that model results may be dependent on the scale of analysis or the definition of underlying reporting units. These are issues originally raised and explored in Openshaw and Taylor (1981) and Tobler (1989), respectively. The important point here is that many analytic methods are known to produce results that could significantly change if the underlying scale or spatial representation changes. For example, if there is correlation found between two variables using Census blocks, the correlation may or may not be found if Census tracks are relied upon to examine these same variables. Tobler (1989) suggests that any analytic method that can be manipulated based on scale or unit definition is frame dependent, and is problematic. Given this, one must seek out and/or develop methods that are not dependent on a particular geographic frame.

To this end, evolution is spatial analytics has in fact been successful to a certain extent in making progress towards addressing aspects of MAUP and frame dependence. For example, Murray (2005) developed a location model where spatial coverage was shown to be less sensitive to scale and/or unit definition. Aldstadt and Getis (2006) developed an endogenous specification of a spatial weights matrix, W, establishing significance for the use of a particular form. Reardon et al. (2008) discuss, review and develop less frame dependent approaches for examining segregation. Of course, research continues along these lines for a number of spatial analytic methods, and more work is necessary.

As noted previously, spatial data often contains some degree of spatial error or uncertainty. Further, methods applied to such data amplifies, propagates and is otherwise impacted by uncertainty/error in different ways. Murray and Grubesic (2012) note and summarize that abstraction, spatial uncertainty (due to geocoding, boundary digitizing, projection, etc.) and varying proximity (e.g., distance measures, neighbor specification, etc.) all co-mingle in unknown ways. The impacts and implications are clearly something that remains to be explored and better understood.

Discussion and Conclusions

An observed trend highlighted in this paper is that there are vast amounts of geographic data, ranging from traditional Census oriented information to GPS tracking of movement and behavior. Individual level data exists and may be readily accessed without the knowledge or consent of individuals. Such data may or may not be beneficial to rural definition or analysis, as this would depend on the intended purpose of a study or inquiry. What is certain, however, is that the wealth of information is generally not them same for rural areas. Thus, many sources of data may not be entirely representative of actual trends and behavior in rural areas given differences in lifestyle but also the role of technology in people's daily activities. There most certainly remains a digital divide between rural and urban areas not to mention different socioeconomic groups. Beyond this, when information does exist for rural areas, the samples are much smaller, raising important significance issues.

This paper has also highlighted that there are a range of spatial analytic methods. GIS is particularly important, offering a range of functions and procedures that enable manipulation of spatial information. Spatial statistical techniques, particularly those included in commercial software packages like ArcGIS and GeoDa, are central to the exploration and assessment of rural conditions and differences. However, there are many other spatial analytic methods, including spatial optimization, geosimulation and ESDA/SDSS. All of these methods continue to evolve in various ways, often because of more data, better insight, increased mathematical specification of spatial detail and better reflecting inherent relationships.

A challenge in the application and utilization of an spatial analytic method is that there are known MAUP and frame dependence issues. Specifically, it is well established that many methods can be sensitive to spatial scale or spatial unit definition. If the scale or unit definition changes, the method may return different results or findings. Recent research has focused on addressing some of these issues, resulting in methods that are more frame independent. However, this is true for only select spatial analytics. In general, there are no definitive ways to deal with MAUP or frame dependence, so this is continuing area of research. As rural definition can vary, certainly this is an issue for any applied spatial analytic technique. In summary then, spatial data can be uncertain, methods applied to data can produce uncertainty and definitions of terms like rural are imprecise and fuzzy. While not explored explicitly, it is clear from the above review that classification matters a lot, and that great care must be taken in formal specification of rural areas.

Computing capabilities and associated software have evolved considerably over the last 20 years. Dealing with big data through the use of GIS and spatial analytics is feasible. The implications for ERS is that rigid definitions linked to a particular spatial scale is not necessary, and that more flexibility should be a feature of future rural data products. Beyond this, an observation is that limited spatial analytics appear to be capable of utilizing rural classification in a sophisticated manner, such as the rural-urban continuum codes. While there has been regression oriented work that accommodates them through the use of dummy variables, it is curious that a broader range of methods seem to be ill-equipped to do so. With the proliferation of open source code and libraries, like PySAL (https://geodacenter.asu.edu/pysal) and CRAN (http://cran.r-project.org/), methods are more accessible. Yet, dealing with a rural coding system based on an ordinal data type requires an extension of most spatial analytic methods. Future research is necessary to explore how a range of methods could be structured to explicitly incorporate rural classification codes along these lines.

References

Aldstadt, J. and A. Getis (2006). "Using AMOEBA to create a spatial weights matrix and identify spatial clusters." Geographical Analysis 38, 327-343.

Anselin, L. (1995). "Local indicators of spatial association—LISA." Geographical Analysis 27, 93-115.

Anselin, L. (2012). “From SpaceStat to CyberGIS twenty years of spatial data analysis software.” International Regional Science Review 35, 131-157.

Anselin, L., A.T. Murray and S.J. Rey (2013). “Spatial analysis.” In The Oxford Handbook of Quantitative Methods, Volume 2, edited by T.D. Little, 154-174 (Oxford, UK: Oxford University Press).

Caschili, S., A. De Montis and D. Trogu (2015). “Accessibility and rurality indicators for regional development.” Computers, Environment and Urban Systems 49, 98-114.

Cayo, M.R. and T.O. Talbot (2003). “Positional error in automated geocoding of residential addresses.” International Journal of Health Geographics 2, 10

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

Fischer, M.M. and A. Getis (eds) (2009). Handbook of Applied Spatial Analysis (Springer: New York).

Getis, A. (2015). “Analytically derived neighborhoods in a rapidly growing West African city: The case of Accra, Ghana”. Habitat international 45, 126-134.

Goodchild, M.F. (2007). “Citizens as sensors: the world of volunteered geography.” GeoJournal 69, 211-221.

Grubesic, T.H., T.C. Matisziw and A.T. Murray (2013). “A strategic approach for improving rural air transport in the United States.” Transport Policy, 30, 117-124.

Grubesic, T.H. and A.T. Murray (2004). “Waiting for broadband: Local competition and the spatial distribution of advanced telecommunication services in the United States.” Growth and Change 35, 139-165.

Grubesic, T. H. and A.T. Murray (2008). “Sex offender residency and spatial equity.” Applied Spatial Analysis and Policy 1, 175-192.

Huang, Q., D.C. Parker, T. Filatova and S. Sun (2014). “A review of urban residential choice models using agent-based modeling.” Environment and Planning B Planning and Design.

Isserman, A M. (2005). “In the national interest: Defining rural and urban correctly in research and public policy.” International Regional Science Review 28, 465-499.

Longley, P.A., M.F. Goodchild, D.J. Maguire and D.W. Rhind (2015). Geographic Information Science and Systems, fourth edition (Wiley: New York).

Murray, A.T. (2010a). “Quantitative geography.” Journal of Regional Science 50, 143-163.

Murray, A.T. (2010b). “Advances in Location Modeling: GIS Linkages and Contributions.” Journal of Geographical Systems 12, 335–354

Murray, A.T. and T.H. Grubesic (2012). “Spatial optimization and geographic uncertainty: implications for sex offender management strategies.” In Community-Based Operations Research, edited by M. Johnson, 121-142 (Springer: New York).

Murray, A.T., T.H. Grubesic and R. Wei (2014). “Spatially significant cluster detection.” Spatial Statistics 10, 103-116.

Murray, A.T., T.H. Grubesic, R. Wei and E.A. Mack (2011). “A hybrid geocoding methodology for spatio-temporal data.” Transactions in GIS 15, 795-809.

Murray, A.T., Y. Liu, S.J. Rey and L. Anselin (2012). “Exploring movement object patterns.” Annals of Regional Science 49, 471-484.

Openshaw, S. (2014). “Geocomputation.” In GeoComputation (second edition), edited by R.J. Abrahart and L.M. See, pp 1-21 (CRC Press: Boca Raton, FL).

Openshaw, S. and P.J. Taylor (1981). “The modifiable areal unit problem.” In Quantitative geography: a British view, edited by N. Wrigley and R. Bennet, 60-69 (Routledge and Kegan Paul: London).

Osth, J., W.A. Clark and B. Malmberg (2014). “Measuring the scale of segregation using k-nearest neighbor aggregates.” Geographical Analysis.

Peng, Z.R. and M.H Tsou (2003). Internet GIS: distributed geographic information services for the internet and wireless networks (Wiley: New York).

Reardon, S.F., S.A. Matthews, D. O’Sullivan, B.A. Lee, G. Firebaugh, C.R. Farrell, and K. Bischoff. (2008). “The geographic scale of metropolitan racial segregation.” Demography 45, 489-514.

Rogerson, P. (2010). Statistical Methods for Geography, 3rd edition (London: Sage)

Shyy, T.-K., I. Azeezullah, I. Azeezullah, R.J. Stimson and A.T. Murray (2014). “Classification for visualizing data: Integrating multiple attributes and space for choropleth display.” In Handbook of Research Methods and Applications in Spatially Integrated Social Science, edited by R.J. Stimson, 265-286 (Edward Elgar: Cheltenham).

Tobler, W.R. (1989). “Frame independent spatial analysis.” In The Accuracy of Spatial Databases, edited by M. Goodchild and S. Gopal, 115-122 (Taylor and Francis: New York).

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

Waldorf, B.S. (2006). “A continuous multi-dimensional measure of rurality: moving beyond threshold measures.” (http://purl.umn.edu/21383)

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

Wong, D.W. and S.L Shaw (2011). “Measuring segregation: an activity space approach.” Journal of Geographical Systems 13, 127-145.

Yao, J. and A.T. Murray (2014). “Locational effectiveness of clinics providing sexual and reproductive health services to women in rural Mozambique.” International Regional Science Review 37, 172-193.