HUMAN GEOGRAPHIES – Journal of Studies and Research in Human Geography
7.2 (2013) 5–13. ISSN-print: 1843–6587/$–see back cover; ISSN-online: 2067–2284–open access
www.humangeographies.org.ro

HUMAN GEOGRAPHIES
Journal of Studies and Research in Human Geography

CONTIGUITY PRINCIPLE FOR GEOGRAPHIC UNITS: EVIDENCE ON THE QUANTITY, DEGREE, AND LOCATION OF PUBLIC USE MICRODATA AREA (PUMA) FRAGMENTATION

Carlos Siordia[a]*, Douglas F. Wunneburger[b]

[a] *Department of Epidemiology at the University of Pittsburgh, USA*
[b] *Department of Landscape Architecture and Urban Planning at Texas A&M University, USA*

Abstract: Social scientists investigating how context varies by geographical location and/or how macro-level phenomenon affects individual outcomes often make use of U.S. Census Bureau Public Use Microdata Sample (PUMS) files where micro-units can only be geographically located to Public Use Microdata Area (PUMA) polygons. Most spatial analysis investigations with PUMAs ignore the fact that many of them are multipart polygons—spatially separated polygons that share the same attribute and are stored as a single feature in a vector file. We briefly discuss the theoretical premises of how geographical boundaries are created for macro units and investigate the quantity, degree, and location of PUMA fragmentation. We argue that the basic contiguity principle (the assumption that spatial analysis uses polygon centroids for solid and contiguous geographic units) in spatial dependence analysis is being violated with many PUMAs in the U.S. mainland—where Texas, California, Tennessee, and Illinois merit special attention. Future research should outline a method for handling multipart polygons in spatial and hierarchical analyses.

Key words: Spatial analysis, Spatial demography, ACS, PUMS, PUMAs, Clustering.

Article Info: *Manuscript Received*: March 12, 2013; *Revised*: November 9, 2013; *Accepted*: November 12, 2013; *Online*: November 18, 2013.

Introduction

A general tenet in spatial demography is that "context" (interchangeably refer to as the environment or the macro-unit) matters and should be accounted for in research (*see* Durkheim, 1951). The interaction between the individual (interchangeably referred to as the micro-unit) and his/her context can operate through social interaction with physically-proximal others and the perception of prevailing norms in immediate-context (*see* Books and Prysby, 1988). The use of geographically referenced ("geo-referenced") information continues to grow in popularity as many research sectors, like public health, begin to expand their use of Geographic Information Systems (GIS: Cromley and McLafferty, 2012). Geo-referencing data has frequently been cited as one of the most important applications of GIS (Mansour et al., 2012) because merging person-level with place-level information may offer researchers the ability to address important questions on how the environment interacts with the individual to affect life chances and health outcomes.

Researchers requiring use of geo-referenced information must either develop their own data (an expensive and time-consuming process) or make use of readily available secondary data. Our project substantively contributes to "place effect" oriented human geographies research by expanding our understanding of how context is measured in a readily available and popular secondary data source. We briefly discuss the theoretical assumptions in geographical units when measuring context. For social sciences, the theoretical importance of environmental measures lies on the largely *implicit* assumption that relationships among geographically distributed micro-level variables vary as a function of macro-level phenomenon and/or geographical location. This view is based on the assumption that everything is related to everything else, but near things are more related than distant things (Tobler, 1970).

Testing the effects and spatial clustering of macro-level phenomenon is important and requires the appropriate delineation of areal boundaries (i.e., geographical boundarization) and measurement of con-

* Corresponding author:
Address: 130 DeSoto Street, Pittsburgh, PA, 15261, USA
Telephone: 1-412-383-1708
Email: cas271@pitt.edu

doi:10.5719/hgeo.2013.72.5

text. Consequently, the boundaries of spatial units matter (*see* Jacquez, 1995). Challenges for estimating "spatial clusters" in human populations dates back several decades (Cuzik and Edwards, 1990) and continues to date with publications highlighting the fact that boundaries in spatial units can encapsulate *single*- and *multi*-part geographies (Siordia and Fox, 2013). Multipart units refer to spatially separated polygons that share the same attribute and are stored as a *single feature* in a vector file. Siordia and Fox (2013) explain that treating multi-part "polygons as contiguous spatial entities erroneously imposes a false structure of contiguity that challenges theoretical and statistical assumptions in geographically aware research" (Pg. 42). Multi-part polygons may be creating "geospatial mismatch" (Siordia and Fox, 2013) by geographically referencing data to the wrong location. Geospatial mismatch has been documented in other publications where probabilistic methods with regression models are suggested for the classification of "mislabeled points" (Mansour et al., 2012). The importance of "positional accuracy" continues to receive extensive coverage in different fields (Hart and Zandbergen, 2013; Shootman et al., 2007; Zandbergern, 2008).

Although frequently absent in print, we posit that there are five principal components underlining the justification for social sciences' ability to derive meaningful geographic boundaries. Potential for measuring structural phenomenon can only exist if macro-measures are: (1) detectable; and made up of geographical polygons that are (2) non-overlapping, (3) contiguous, (4) non-porous, and (5) non-fluid. For example, measuring "neighborhood economic deprivation" (Burdette, 2013; Law and Quick, 2013) requires the ability to conceptualize, operationalize, and develop data that capture the concept. If the ability for measuring the macro-measure is present, investigators must then spatially reference the measure by using *solid* polygons that are mutually exclusive. "Solid polygons" refers to areas where geographical units do not have gaps, constantly shift, or are fragmented. "Mutually exclusive" refers to geographical units that do not non-overlap in space. The main point here is that the "the contiguity of the geographical polygon is an implicit *and* necessary condition" (Siordia and Fox, 2013, pg. 44) in spatial analysis seeking to georeference data (Grubesic and Matisziw 2006). Our discussion focuses on the contiguity principle.

Social scientists' ability to capture the effects of context on micro-outcomes and the spatial autocorrelation within such measures depends on their ability to avoid violating the postulates outlined above. Spatial autocorrelation refers to the fact that "measurements made at nearby locations may be closer in value than measurements made at locations farther apart" (Srinivasan and Venkatesan, 2013, Pg. 1). If the contiguity principle proofs necessary, then the social significance, reliably, and validly of macro-measures will be determined by our ability to properly *demarcate boundaries*. The current study only focuses on a particular challenge (i.e., polygon fragmentation—the presence of multi-part polygons) when attempting to demarcate geographical boundaries.

Spatial Reference and Dependence

Geographically referenced demographic data has many analytical uses. Hierarchical (Raudenbush et al, 2002) and spatial approaches like local clustering (Anselin, 1995) and geographically weighted regression (Fotheringham et al 2002) have been applied across many investigation topics (e.g., Fullerton, 2012; Liu and Painter, 2012; Yang and Matthews, 2012). At the core of these investigative approaches is the idea that context plays a significant role in an individual's behavior—that "place maters" (Laraia et al., 2012).

There are five general uses of spatially reference demographic data: (1) to create customized macro-measures in a spatial clustering investigation (Wang, 2007); (2) where the micro-unit is the unit of analysis in a model that includes macro-unit measures (e.g., Raphael and Stoll, 2010); (3) where the micro-unit of analysis and its geographical location is required in the estimation of the model (e.g., Barrios et al, 2010); (4) where the macro-unit, as the unit of analysis, requires a customized estimate (e.g., Yu and Myers, 2007); and (5) where the geographic polygon is used in areal interpolation procedures (Salvatore et al, 2007). Polygon fragmentation could pose problems for each of these as follows: (1) fallible polygon centroids; (2) indeterminable inaccuracy of macro-level measures assigned to micro-units across fragments; (3) variation in precision for micro-units' approximate physical location; (4) misleading aggregations; and (5) ambiguous areal interpolations from fragments (for more details *see* Siordia and Fox, 2012). Our investigation only focuses on the first challenge: showing evidence that multipart polygons exist. The crucial point is that "polygon centroids" in multi-part geographical units may be calculated outside the polygon—thereby creating a geo-spatial mismatch.

Because a polygon's demographic attribute (e.g., percent who do not speak English) is partially a function of the same attribute in neighboring units (*see* Flint et al, 2000), investigating spatial dependence can compliment cross-level modeling (*see* LeSage and Pace, 2009). Explorations on spatial autocorrelation are fundamentally interested in knowing if/how contexts vary as a function of location, in capturing how demographic attributes can be spatially nonstationary (i.e. vary as a function of geographical space). Others have explained elsewhere that ignoring spatial dependence has theoretical and statistical implications (Vilalta, 2011). Thus, when possible, researchers

should seek to investigate how spatial non-stationary processes play a role in macro-level measures.

Measurements of spatial dependence were developed many years ago. The general approach was based on the strong statistical assumptions that deviations in observed point patterns can be detected by comparing them to a condition where random point-process are stationary (non-moving) and isotropic (maintains degree of movement in all directions) (Baddeley and Silverman, 1984). Ripley (1976; 1977; 1981) formally introduced the *K* function (used in spatial clustering analysis) and proved its reliability as a statistical tool to analyze second order moment in a point pattern process. The K function can be given as: $K(h) = \{E[N_o(h)]\} \div \lambda$, where the numerator is the expected number of events lying within distance *h* in an arbitrary event of process, and where the denominator λ is the intensity of the process (*also see* Diggle, 1983). So that if we are said to have a *benchmark* point pattern process is $K(h) = \pi h^2$, then $K(h) < \pi h^2$ signals a "regular" point in the pattern, while $K(h) > \pi h^2$ indicates a "clustered" process (Galvis et al, 2009)—where the intensity in movement is over the threshold.

Implicit here is the statistical assumption that benchmark point pattern processes are measured between points representing singlepart spatial units. More simply, spatial clustering techniques may statistically be assuming "movements" occur over continuous space (Jones and Casetti, 1992) where "local instabilities" are clusters (Openshaw, 1993). For example, measurements of local spatial association (Anselin, 1995) make use of the k function by observing the spatial autocorrelation (the correlation of one variable with itself as a function of location). When defining local indicator of spatial association (LISA) statistics, Anselin (1995) explains that local spatial dependence is present when similar values significantly cluster. With similar approaches using points (below referred to as polygon centroids), analysis of mapped planar point patterns then focuses on the behavior of point-attributes and their distances between pairs of points in the pattern (Baddeley and Silverman, 1984). Here again, we have the implicit idea that a solid and singlepart spatial unit is compared with similar spatial entities in a given neighborhood-bandwidth.

Polygon Centroids

Most statistical approaches measuring spatial nonstationarity are determined based upon a polygon's *geometric centroid*. Centroids use *planimetric* calculations—make use of spatial references from projected (rather than spherical or geodesic) space. Most centroids represent a polygon's mean center based on the weighted average of its *x*- and *y*-geographic coordinates (Mitchell, 2005). A centroid is in essence the center of gravity on which the polygon can be balanced and it is a common way of summarizing the location of an attribute in spatial analysis (Goodman et al., 2012; Wang, 2010). In most research, the distance between spatial units is measured using these feature centroids. This is most appropriate where the polygons are roughly the same size and shape (Mitchell, 2005)—and where polygons are made up of a singlepart shape where reaching all the points within the polygon can be done without ever stepping out of the spatial unit.

Since investigating spatial nonstationarity is crucial, accounting for it while using secondary data sources requires that scientists make use of whatever geographic polygons are available with the data. The use of multipart geographical units available in secondary data sources presents a challenge with a spatial dimension. For example, detail-rich Public Use Microdata Sample (PUMS) files (more details below) only allow individual-level units to be spatially referenced to Public Use Microdata Area (PUMA) polygons. Although many researchers may be unaware of it, a substantial amount of PUMAs are made up of *multipart* polygons. We believe burgeoning theories on spatial dependence and the methodologies employed in its measurement must pay careful attention to such a special condition. Our paper makes a substantive contribution to the literature by highlighting this challenge.

A discussion on Figure 1 may help understand the implications of using geometric centroids with singlepart and multipart polygons. On the left, in Figure 1, we have a singlepart polygon and on the right (enclosed in the dotted line) we have a multipart polygon, where the dark circle for both represents their geometric center. With the singlepart polygon, the centroid is within the spatial unit and adequately represents the center of the unit. In contrast, the centroid in the multipart polygon is outside any of the fragments and is less representative of the *center* represented in the singlepart polygon. Please note that the fragments that constitute the multipart-polygon vary in shape, size, and distance from each other. Although not shown in Figure 1, the space between the multipart-polygon fragments would, in the case of PUMAs, be filled by other PUMA units. For an example of an actual map displaying multipart polygons *see* Siordia and Fox (2013).

Our core theoretical question is: Is polygon *contiguity* a necessary condition when investigating spatial nonstationarity? Contiguity refers to geographical units sharing common boundaries (as is the case with singlepart polygons) while *noncontiguity* refers to spatial units made up of multiple parts that do not share a boundary (as is the case with multipart polygons). A person travelling to all the internal points in a contiguous spatial unit would not have to leave the polygon, while a person travelling in a noncontiguous polygon would have to exit the geometrical unit at

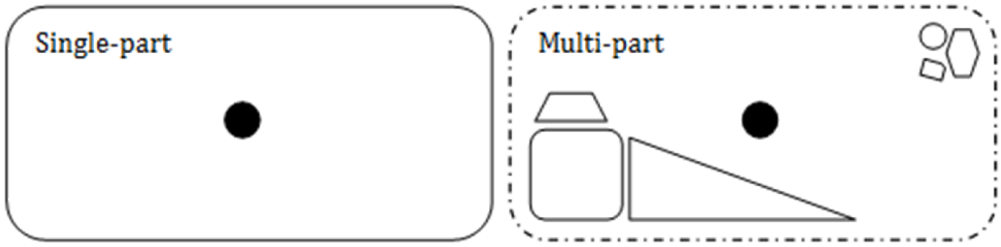

Figure 1. Single-part and multi-part polygons

some point to enter it again in a different location (Cova and Church, 2000). A *fragmented* polygon then refers to a multipart spatial unit. If polygon contiguity is a necessary condition for investigating spatial nonstationarity, and we contend this is the case, then multipart polygons pose a problem.

Our core argument is that polygon contiguity is *necessary* in investigations seeking to account for spatial nonstationarity. Exploring how macro-level measures interact with micro-level outcomes and how statistical relationships may vary as a function of geographical location requires, given theoretical and statistical assumptions, that geographical attributes be derived from contiguous polygons. Polygon non-contiguity can have a significant impact on spatial analysis and theory (Grubesic and Matisziw, 2006), because the treatment of multipart polygons as contiguous units imposes a false structure that may produce systemic errors in theory and methods. Because investigations on spatial nonstationarity (using the center of the polygon to measure location) are theoretically premised on the assertion that a polygon's attribute is constituted by a contiguous spatial unit, the presence of PUMA fragmentation merits research attention.

Specific Aims

Our research question is: *Is PUMA polygon discontiguity present in US mainland states*? The specific aims of this paper are to investigate the quantity and degree of fragmentation at the state-level across US contiguous states and the District of Columbia. Our project quantifies and spatially references where the basic *contiguity requirement* is violated with PUMA polygons. In closing, we discuss implications of the findings, limitations, and what future research should be undertaken.

Materials and Methods

In the United States (US), the use of US Census Bureau microdata (i.e., individual-level data) aids the allocation of governmental funds and services (see Reamer 2010). In order to provide non-Census researchers the opportunity to work with information-rich individual-level data, the Census releases Public Use Microdata Sample (PUMS) files. In order to ensure the confidentiality of survey respondents, they undertake several procedures (e.g., capping age maximum at 99) and only allow the public data users the ability to geographically locate individuals to spatial units referred to as Public Use Microdata Areas (PUMAs). Although our investigation explores the geometrical attributes of PUMAs and discusses their implications for spatial analysis, similar analysis could be done with other spatial entities.

The US Census Bureau releases PUMA geographies using a combination of numeric or alphanumeric codes to spatially reference micro-units and where each spatial unit must contain at least 100,000 people. The geographical boundaries of PUMAs are created from collaboration between State Data Centers (SDCs) and the federal goverment (i.e., US Census Bureau). Criterion for PUMA delineation has changed since 1990 when this approach was first used (Siordia and Fox 2012).We investigate the quantity and location of multipart PUMAs (across mainland states) by using Topological Integrated Geographic Encoding Referencing (TIGER) Shapefiles (Zandbergen et al., 2011).

Geographic features can be represented digitally for application in geographic information systems using any of a number of possible geospatial vector formats. For the purposes of this study, all analysis employs geospatial features in shapefile format (ESRI 1998). Our 2007 TIGER/Line PUMA shapefiles contain the geographic boundaries as of January 1, 2007, which includes a Census 2000 vintage geography. In 2000, counties, minor civil divisions, incorporated places, and tracts were used as the *building blocks* to deliniate the geographic boundaries of PUMA polygons. PUMA multipart data is produced using ArcGIS® 9.3 [software by ESRI. ArcGIS® and ArcMap™ are the intellectual property of ESRI and are used herein under license (Copyright © ESRI, all rights reserved) for more information about ESRI® software, please visit www.esri.com] (ESRI 2011).

We use the ArcGIS 9.3 *Explode* tool in advanced editing to identify the quantity and location of multipart polygons. From the produced shapefile containing the single-part polygons (which retain their original PUMA identification number along with a new polygon id) we generated our analytic PUMA-polygon sample.

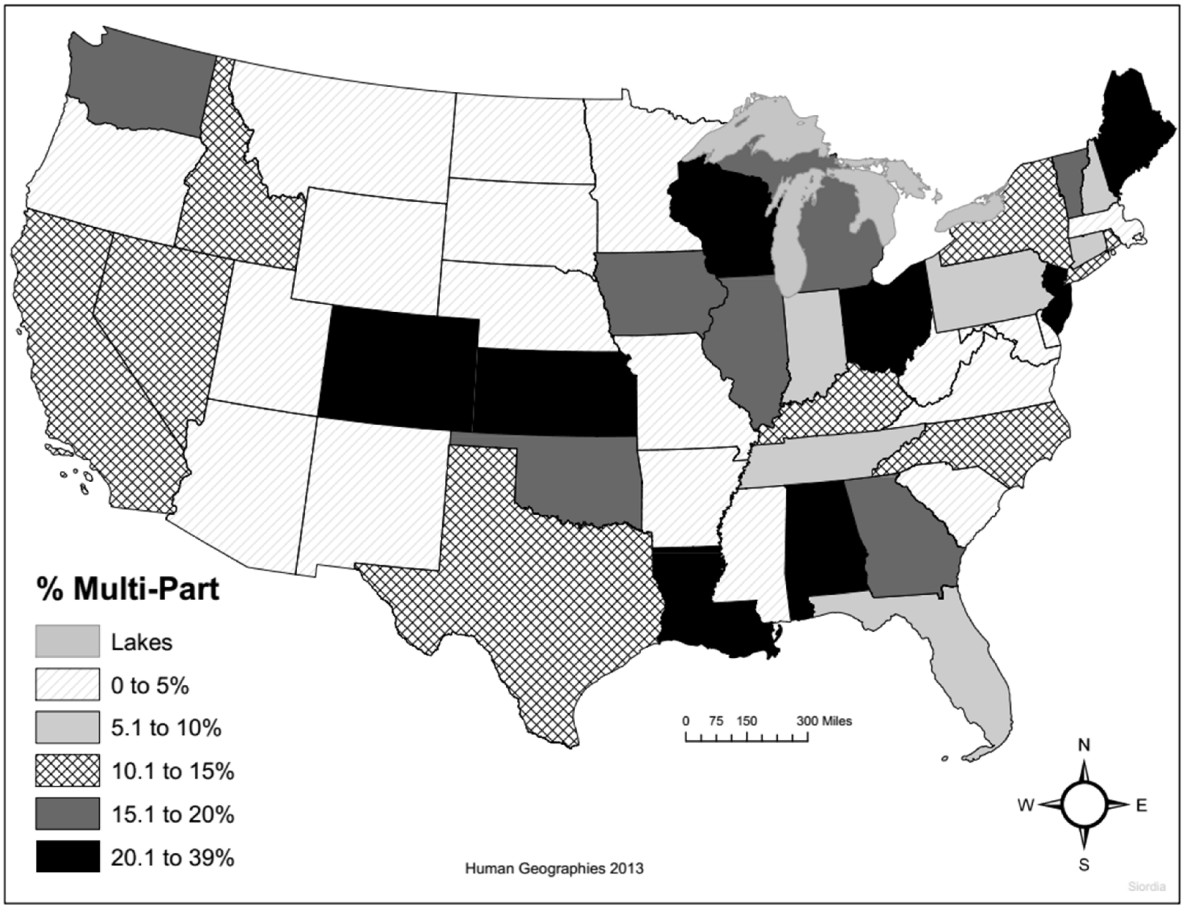

Figure 2. Concentration of PUMA Fragmentation by State

The quantification and localization of fragmentation, from the resultant data output in the previous step, is done by importing our data and managing it in SAS 9.2 (Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA).

Our descriptive analysis explores PUMA fragmentation by state. We computed a state's *percent* of PUMA fragmentation with the following equation: [(total number of PUMAs ÷ number of fragmented PUMAs)*100]. This measure captures the amount of PUMA fragmentation by state. We also calculated a ratio to measure a state's *degree* of PUMA fragmentation by using the following equation: (number of fragments from multipart PUMAs ÷ number of fragmented PUMAs). This measure assesses the degree to which multipart PUMAs are fragmented.

Results and Discussion

From the PUMAs (i.e., total number of PUMAs in state) in Table 1, we saw the following states had more than 100 PUMA polygons (number given in parentheses): California (233); Texas (153); New York (143); and Florida (127). Our table provided the measures for states where at least one PUMA fragmentation was present. The following states were omitted from the table because they have zero fragmentations (number of PUMAs in state given in parentheses): Delaware (6), Mississippi (23); Montana (7); North Dakota (5); Nebraska (14); Utah (16); West Virginia (12); Wyoming (4); and (5) the District of Columbia. Please note that although these PUMA non-fragmenting states had a small count of polygons, some of the fragmenting states in our tables had fewer PUMA counts (e.g., Vermont, Maine, and Idaho). These states are in the "0 to 5%" category in Figure 2.

Many factors may be at play in how PUMAs become fragmented—since they are the product of federal and local government efforts (*see* Siordia and Fox 2012). Even though PUMAs are driven by population size, and are shaped by their building blocks (e.g., blocks, tracts) the amount of rural space in the state does not seem to play a role in the presence of fragmented PUMAs.

Although an internal investigation on the procedures involved (between the federal and local government) in the delineation of PUMA geographical boundaries would be required to understand why some states avoided fragmenting PUMAs, non-

fragmenting states should be commended for avoiding (whether intentionally or non-intentionally) the formation of multipart polygons.

We now evaluate states' percent of fragmentation and then discuss the degree to which multipart polygons are fragmented for each state.

Relative percent of fragmentation

From Table 1, we first dealt with the "% Fragmented" column. Although the table is sorted (from largest to smallest) on the degree of fragmentation, we made note of states where at least 20% of the PUMAs were fragmented. Figure 2 also displays the data for the US states. The State with the most geometrical fragments was Oklahoma with 7 of its 18 PUMAs being fragmented. It is followed by Nevada and North Carolina with one third of their PUMAs fragmented. Texas has 41 of its 153 PUMAs fragmented while California has 46 of its 233 PUMAs fragmented. The other states (VT, KS, WI, ID, ME) with at least 20% of their PUMAs fragmented have a smaller PUMA counts and contain between a 20% and 25% amount of multipart PUMA polygons. From our descriptive analysis on the states' *percent* of PUMA fragmentation, we found that the small PUMA count states of Oklahoma and Nevada had the largest percent of fragmentation, while the large PUMA count states of Texas and California had 27% and 20% multipart PUMA polygons respectively.

Degree of fragmentation

We now turn our attention to the degree to which multipart PUMAs are fragmented. From Table 1, we see that Tennessee and South Carolina have a ratio of about 24 fragments for each of their multipart PUMAs. Illinois follows them with a ratio of 22, while Michigan and Alabama have about 15 fragments for each of their multipart PUMAs. North Carolina and Kansas have a 12 and 11 ratio respectively. Texas and California have about 7 and 6 fragments for each of their many multipart PUMAs accordingly.

From this descriptive analysis on the states' *degree* of PUMA fragmentation, we find that Tennessee and Illinois have the largest degree of multipart PUMA fragmentation within moderate PUMA count states. Because of their large PUMA count, it is worth mentioning that Texas and California on average split their multipart PUMAs into six fragments. Please note that in contrast, the large PUMA count states of Florida and New York only have, on average, 3 fragments for their few multipart polygons—this too is a commendable accomplishment.

Conclusions

In answer to our research question, we found that PUMA polygon discontiguity was present in US mainland. The basic contiguity requirement for spatial analysis is being violated with many PUMAs in 40 out of 49 mainland US states. Texas and California are of the biggest concern within *large* PUMA count states, Tennessee and Illinois within *moderate* PUMA count states, while Oklahoma and Nevada merit attention even though they have a *small* PUMA count. The fact that PUMA fragments are dispersed over a sea of complete- and fractured-PUMA polygons may allow the "inappropriate" overlapping of feature centroids, given the geographical proximity of the flawed centroids from fragments, which could cause the detection of a false-positive (saying that the a relationship is statistically significant when it is not) when investigating spatial dependency. When researchers are exploring spatial clustering or modeling spatial autocorrelation, they should insure their geographical units are not made up by multi-part polygons.

One limitation in our study is that we do not provide the reader with information on the size and shape of fragments or the encaptured square-mile area within multipart polygons. For example, some fragments may be made up of one square-mile while other may be constituted by a polygon ten times that size. Some may have more homogenous shape configurations where others are made up of widely disperse slithers. Also, fragments may be dispersed over a one-mile radius or a 10-mile radius. Out investigations does not offer insights to these important PUMA fragmentation elements. Future research should pursue these questions. Social scientist should also seek to explore if research in other fields on "edge detection" (e.g., Safner et al, 2011) can help formulate more theory driven macro-level measures and Bayesian clustering techniques. By employing such approaches, the implicit assumption that macro-level boundaries reliably and validly demarcate where an abrupt change in how context occurs can be investigated.

This paper, notwithstanding the limitations, makes a substantive contribution by highlighting the presence, quantity, degree, and location of PUMA discontiguity by states in the US mainland. Building a bridge between geographically aware research and social science endeavors, requires that we expand our dialogue on how our rapidly evolving methods and software can reliably and validly investigate (or fail to do so because of multipart polygons) our research questions. Because there may be many social and policy consequences from deriving erroneous conclusions through flawed approaches, more research on the implications of polygon fragmentation in the analysis of spatial nonstationarity is necessary.

Table 1. Count of total, non-fragmented, and fragmented PUMAS sorted by degree of fragmentation

State	PUMAs	Non-Fragmented	Fragmented	%Fragmented	Fragments	Degree
OK	18	11	7	15.9%	179	25.6
NV	15	10	5	14.8%	98	24.5
NC	58	41	17	10.3%	196	21.8
TX	153	112	41	14.7%	152	15.2
VT	4	3	1	16.7%	74	14.8
KS	21	16	5	29.3%	207	12.2
WI	31	24	7	23.8%	54	10.8
ID	9	7	2	11.1%	38	9.5
ME	10	8	2	22.2%	18	9
CA	233	187	46	14.3%	9	9
AL	30	25	5	26.8%	300	7.3
TN	44	37	7	9.8%	29	7.3
CO	38	32	6	38.9%	48	6.9
IA	19	16	3	19.7%	286	6.2
OH	91	77	14	22.6%	34	4.9
SC	27	23	4	4.8%	14	4.7
MI	68	58	10	15.8%	26	4.3
RI	7	6	1	10.5%	8	4
SD	7	6	1	3.7%	4	4
LA	36	32	4	33.3%	19	3.8
WA	46	41	5	15.8%	11	3.7
AR	19	17	2	4.2%	7	3.5
IL	87	78	9	15.4%	46	3.3
KY	30	27	3	10.9%	16	3.2
MO	41	37	4	3.9%	16	3.2
NH	11	10	1	6.7%	3	3
MN	37	34	3	1.9%	3	3
NM	15	14	1	4.2%	16	2.7
PA	92	86	6	6.5%	13	2.2
NJ	61	58	3	25.0%	2	2
GA	63	60	3	20.0%	4	2
NY	143	137	6	14.3%	2	2
IN	48	46	2	10.0%	6	2
CT	25	24	1	9.1%	2	2
FL	127	122	5	8.1%	6	2
OR	27	26	1	4.9%	6	2
AZ	36	35	1	4.0%	2	2
VA	42	41	1	2.8%	2	2
MD	44	43	1	2.4%	2	2
MA	52	51	1	2.3%	2	2

References

Anselin, L 1995,'Local indicators of spatial association-LISA', *Geographical Analysis*, 27, pp. 93-115.

Baddeley, AJ & Silverman, BW 1984,'A Cautionary Example on the Use of Second-Order Methods for Analyzing Point Patterns', *Biometrics*, 40, pp. 1089-1093.

Barrios, T, Diamond, R, Imbens, GW & Kolesar, M 2010, Clustering, Spatial Correlations and Randomization Inference. National Bureau of Economic Research, NBER Working Paper No. 1576.

Books, J & Prysby, C 1988, 'Studying contextual effects on political behavior: a research inventory and agenda', *American Politics Quarterly*, 16, p. 211-238.

Burdette, AM 2013, 'Neighborhood context and breastfeeding behaviors among urban mothers', *Journal of Human Lactation*, 29, pp. 597-604.

Cova, TJ & Church, RL 2000,'Contiguity constraints for single-region site search problems', *Geographical Analysis*, 32, pp. 306-329.

Cromley, EK & McLafferty, S 2012, *GIS and public health,* Guilford Press.

Cuzick, J & Edvards, R 1990,'Spatial clustering for inhomogeneous populations', *Journal of the Royal Statistical Society*, 52, pp. 73-104.

Diggle, PJ 1983, *Statistical analysis of spatial point patterns*, Academic Press, London.

Durkheim, É 1951, *Suicide: a study in sociology*, Free Press.

ESRI 1998, Shapefile Technical Description: An ESRI White Paper, Redlands, California.

ESRI 2011, ArcGIS Desktop, Release 10, Environmental Systems Research Institute, Redlands, CA.

Flint, C, Harrower, M & Edsall, R 2000, But how does place matter? using Bayesian networks to explore a structural definition of place. Paper presented at the New Methodologies of the Social Sciences Conference, University of Colorado Boulder.

Fotheringham, AS, Brunsdon, C & Charlton, ME 2002, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley, West Sussex, UK.

Fullerton, AS 2012,'Spatial agglomeration and wages in the U.S. biotechnology sector', *Sociological Spectrum*, 32, pp. 61-80.

Galvis, L, Guertin, PJ & Meyer, WD 2009, *Actionable cultural understanding for support to tactical operations: the effect of data quality on spatial analysis results. Report*, ERDC/CERL TR-09-15.

Goodman, JM, Owens, PR & Libohova, Z 2012,'Predicting soil organic carbon using mixed conceptual and geostatistical models', *Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping*, , 10-13 April 2012, CRC Press, Sydney, Australia.

Grubesic, TH & Matisziw, TC 2006,'On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data', *International Journal of Health Geographics*, 5, pp. 1-15.

Hart, TC & Zandbergen, PA 2013,'Reference data and geocoding quality: Examining completeness and positional accuracy of street geocoded crime incidents', *Policing: An International Journal of Police Strategies & Management*, 36, pp. 263-294.

Jacquez, GM 1995,'The map comparison problem: tests for the overlap of geographic boundaries', *Statistics in Medicine*, 14, pp. 2343-2361.

Jones, JP & Caselli, E 1992, *Applications of the expansion method*, Routledge, London.

Laraia, BA, Karter, AJ, Warton, EM, et al. 2012,'Place matters: neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of Northern California (DISTANCE)', *Social Science Medicine*, 74, pp. 1082-1090.

Law, J & Quick, M 2013,'Exploring links between juvenile offenders and social disorganization at a large map scale: a Bayesian spatial modeling approach', *Journal of Geographical Systems*, 15, pp. 89-113.

LeSage, JP & Pace, RK 2009, *Introduction to spatial econometrics*, CRC Press, Boca Raton.

Liu, CY & Painter, G 2012,'Travel behaviour among Latino immigrants: the role of ethnic concentration and ethnic employment', *Journal of Planning Education and Research*, 32, pp. 62-80.

Mansour, S, Martin, D & Wright, J 2012,'Problems of spatial linkage of a geo-referenced Demographic and Health Survey (DHS) dataset to a population census: A case study of Egypt', *Computers, Environment and Urban Systems*, 36, pp. 350-358.

Mitchell, A 2005, *The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns and Relationships and Zeroing In: Geographic Information Systems at Work in the Community,* ESRI Press, US.

Openshaw, S 1993,'Some suggestions concerning the development of artificial intelligence tools for spatial modelling and analysis in GIS' in MM Fischer & P Nijkamp (eds), *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, pp. 17-33, Springer Verlag, Berlin.

Raphael, S & Stoll, MA 2010,'Job sprawl and the suburbanization of poverty. Metropolitan Policy Program at Brookings', *Metropolitan Opportunity Series*, March: 1-21.

Raudenbush, SW & Bryk, AS 2002, *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn, Thousand Oaks, Sage Publications, California.

Reamer, AD 2010, *Surveying for dollars: the role of the American Community Survey in the geographic distribution of federal funds,* Metropolitan Policy Program at Brookings, Washington D.C.

Ripley, B 1976,'The second-order analysis of stationary point processes', *Journal of Applied Probability,* 13, pp. 255-266.

Ripley, BD 1977,'Modelling spatial patterns (with discussion)', *Journal of the Royal Statistical Society*, Series B 39, pp. 172-212.

Ripley, BD 1981, *Spatial statistics*, Wiley, New York.

Safner, T, Miller, MP, McRae, BH, Fortin, M & Manel, S 2011,'Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics', *International Journal of Molecular Sciences*, 12, pp. 865-889.

Salvatore, S, Chavers, JM, Nixon, LC & McQuiddy, MR 2007,'From here to there: methods of allocating data between census geography and socially meaningful areas', *Social Science Research*, 36, pp. 897-920.

Siordia, C & Fox, A 2013, Public Use Microdata Area fragmentation: research and policy implications of polygon discontiguity, Spatial Demography, 1(1), pp. 42-56.

Schootman, M, Sterling, DA, Struthers, J et al. 2007,'Positional accuracy and geographic bias of four methods of geocoding in epidemiologic research', *Annals of epidemiology*, 17, pp. 464-470.

Srinivasan, R & Venkatesan, P 2013,'Bayesian model for spatial dependance and prediction of tuberculosis', *International Journal*, 3, pp. 2307-2083.

Tobler, W 1970,'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography*, 46, pp. 234-240.

Vilalta, CJ 2011,'The spatial dependence of judicial data', *Applied Spatial Analysis and Policy*, pp. 1-17.

Wang, F 2010, *Quantitative methods and applications in GIS,* CRC Press.

Wang, Q 2007,'Linking home to work: ethnic labor market concentration in the San Francisco consolidated metropolitan area', *Urban Geography*, 27, pp. 72-92.

Yang, T & Matthews, SA 2012,'Understanding the non-stationary associations between distrust of the health care system, health conditions, and self-rated health in the elderly: a geographically weighted regression approach', *Health and Place*, 18, pp. 576-585.

Yu, Z & Myers, D 2007,'Convergence or divergence in Los Angeles: three distinctive patterns of immigrant residential assimilation', *Social Science Research*, 36, pp. 254-285.

Zandbergen, PA 2008,'Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy', *Transactions in GIS*, 12, pp. 103-130.

Zandbergen, PA, Ignizio, DA & Lenzer, KE 2011,'Positional accuracy of TIGER 2000 and 2009 road networks', *Transactions in GIS*, 15, pp. 495-519.