Spatial Demography
Series Editors: Frank M. Howell · Jeremy R. Porter

Frank M. Howell
Jeremy R. Porter
Stephen A. Matthews *Editors*

Recapturing Space: New Middle-Range Theory in Spatial Demography

Springer

Spatial Demography Book Series

Volume 1

This series explores the application of spatial science to demographic information, including the methods, theory, and applications to societal problems. It captures the rapidly expanding knowledge-base of how human behavior and social processes are both shaped by place and time as well as change localities themselves over time. The books in this series examine both individuals as well as institutions and include all methodological, geographical, and topical research or treatments of the interaction between societies and place. It encompasses racial segregation, crime, urban sprawl, agricultural production, engagement in community life, migration, commuting, business location, technology, environmental quality, elections to public office, and other important societal phenomena. This series adopts an international and interdisciplinary approach in its detail of the tools, techniques, and theoretical perspectives used in spatial demography. The individual volumes will help demographers better understand when, how, and why space matters in social behavior and institutions.

More information about this series at http://www.springer.com/series/13093

Frank M. Howell • Jeremy R. Porter
Stephen A. Matthews
Editors

Recapturing Space: New Middle-Range Theory in Spatial Demography

Springer

Editors
Frank M. Howell
Department of Sociology
Mississippi State University
Ridgeland, MS, USA

Jeremy R. Porter
Brooklyn College and Graduate Center
City University of New York
Brooklyn, NY, USA

Stephen A. Matthews
Department of Sociology &
Department of Anthropology
State College, PA, USA

Spatial Demography Book Series
ISBN 978-3-319-22809-9 ISBN 978-3-319-22810-5 (eBook)
DOI 10.1007/978-3-319-22810-5

Library of Congress Control Number: 2015954961

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Recapturing Spatial Approaches to Social Science Problems

Frank M. Howell, Jeremy R. Porter, and Stephen A. Matthews

1.1 Introduction

In *Geographical Sociology*, Porter and Howell (2012) describe a minimal set of criteria that a spatial informed analysis should contain. Specifically they argued that spatial analysis should be spatial in:

(i) *middle-range theoretical framework*;
(ii) one or more key *concepts*;
(iii) the *operationalization* of concepts; and,
(iv) the *analytical* methods used to explore or test the theoretical framework.

The integration of these four elements within a research project would bring together spatial theory and empirical research—or middle-range theory, *a la* Robert Merton (1968)—and in doing so help advance spatial theory in disciplines such as sociology, demography and across the social sciences. Advancing spatial theory is a critical need as the ready availability of geospatial data and the refinement and emergence of analytical tools—geographic information systems (GIS), spatial analysis, and spatial statistics—has not been accompanied by parallel theoretical development (Matthews et al. 2012).

F.M. Howell
Department of Sociology, Mississippi State University, 416 Bay Cove,
Ridgeland, MS 39157, USA
e-mail: frankmhowell@gmail.com

J.R. Porter (✉)
Brooklyn College and Graduate Center, City University of New York, Brooklyn, NY, USA
e-mail: jporter@brooklyn.cuny.edu

S.A. Matthews
Department of Sociology & Department of Anthropology, State College, PA, USA
e-mail: matthews@psu.edu; sxm27@psu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_1

In conceiving this edited collection we sought contributions from scholars representing a diverse set of disciplines, that could help illustrate the approach, results, and testing of middle-range theory across the social sciences. The chapters that follow are grouped in to three main *Parts* broadly covering (*Part I*) theory, (*Part II*) concepts and measures, and (*Part III*) research practice. In addition, the book concludes with a final set of chapters (*Part IV*) focusing on instruction in the area of spatial analysis and concluding remarks.

Part I: Theory contains five chapters (Chaps. 2, 3, 4, 5 and 6). While spatial analysis is "on a roll," **Logan** in *Challenges of Spatial Thinking* (Chap. 2) reminds the social scientist to pay attention to theoretical and substantive concerns, and that these should guide our use of the new spatial tools. To Logan, spatial thinking is about "where things are or where they happen, and it is especially about *where they are in relation to others*." In this context the concepts that are critical to spatial thinking—distance, proximity, exposure, and access—are all rooted in relative locations. Distance and spatial dependence are often the starting point for the development of theory, creation of measures, and determination of which analytical techniques to use. Logan's chapter includes a review the contributions of other scholars in sociology that have used geospatial data and analytical methods; weaving in examples from his own research too. Logan reminds us that the most effective use of new data and tools requires a greater emphasis on spatial thinking and sensitivity in the use of spatial concepts.

While Logan emphasized *space*, **Siordia and Matthews** argue that new scholarship in demography requires a synthesis of existing theories and conceptualizations of *place*. Similar to Logan, in *Extending the Boundaries of Place* (Chap. 3), they argue that more rigorous conceptual models will help enhance understanding of the processes by which place 'gets into people.' Siorda and Matthews suggest that for the most part, studies of the relationship between demographic and health outcomes and place have been based on several conventional and naïve assumptions about place. Specifically, a *discrete* view of the world is reinforced by geographically-based contexts and data structures but these ignore the normal day-to-day activity spaces traced out by people as they navigate their complex lives in *continuous* space. Siorda and Matthews argue for more clear thinking on the processes and mechanisms linking people to place. They identify the need for the use of units of analysis that reflect the spatial and temporal scales of human behavior and they suggest that there are several geographical contexts—based on functional ties—that would appear to be more relevant than others in shaping micro-level behaviors.

The chapter by **Brazil** reviews the recent history of neighborhood effect research and some of the methodological issues that plague this kind of research. Brazil noted that while neighborhoods seem to matter current understanding of the mechanisms as to why they matter remains unclear. In *Putting the "Place" Back into Neighborhood Effects Research: Using Place-based vs. Person-based Interventions to Measure Neighborhood Effects* (Chap. 4), Brazil develops a framework that allows researchers to compare the efficacy of people-based and place-based interventions through the decomposition of the total intervention effect into natural direct and indirect effects. Public policy generally has considerably less ability to influence individual behavior than to affect neighborhood quality. However, as Brazil notes, place-based interventions may provide a more suitable testing grounds for middle-range theories and minimize some of the interpretability issues that hinder person-based interventions.

Chapters 5 and 6 include contributions focusing more explicitly on the use of geospatial data and spatial concepts in demography in US and international research, respectively. **Wong** has been a leader in race/ethnic segregation research, and specifically in the development of new methods for measuring segregation. In *From Aspatial to Spatial, from Global to Local and Individual: Are we on the Right Track to Spatialize Segregation Measures?* (Chap. 5), Wong reflects on how space was (re)introduced into measuring segregation and how available census-type data can both facilitate and constrain how we measure segregation. Specifically, Wong notes that while existing measures are spatial in nature, it is not clear whether they capture the basic notions of segregation, and the spatial dimensions of segregation. In a wide-ranging chapter, Wong reviews and critiques existing practice and speculates on directions, challenges, and potential new measures in measuring segregation in light of the increasing use of individual-level and survey-based data. This chapter illustrates the value of seeking alignment among concepts, data, and methods.

Turning to international demography, *Demography is an Inherently Spatial Science* (Chap. 6) is an update and synthesis of some of **Weeks'** own writings on the interconnectedness of demographic processes and different social transitions that constitute demographic transition, on spatial demography, and on the integration of theory, data and method to examine spatial inequalities in urban West Africa (Weeks 2004, 2011; Weeks et al. 2013). To Weeks, demography is in the process of evolving from a spatially aware science to a spatially analytic science. Demographers are increasingly aware of the spatial nature of demographic transitions and the three spatial elements—space, place, and scale—related to the timing and pattern of these transitions. The second half of Week's chapter is focuses on the study of fertility in Ghana and provides an exemplar illustration of a research project informed by spatial concepts coupled with the use of spatial statistical methods. We close *Part I* with Weeks' chapter as we see it as a bridge to the chapters on concepts and measures (*Part II*).

Part II: Concepts and Measures contains two chapters focused on the methodologically-centered synthesis of disparate literatures on spatial connectivity and its application in an interdisciplinary contexts. **Mobley and Bazzoli** are health economists, and in their chapter on *Modeling 'Dependence of Relevant Alternatives' in Consumer Choice: A Synthesis From Disparate Literatures* (Chap. 7) they show that incorporating spatial dimensions in to hospital choice research can enhance model tractability and plausibility. They review the literature on hospital choice and find many studies using inappropriate 'independence of irrelevant alternatives' (IIA) models, and only a few using the 'dependence of relevant alternatives' (DRA) formulation. Mobley and Bazzoli compare these methods and show that tractable DRA models exist for situations where location matters; e.g., modeling healthcare provider choice in modern urban markets. The authors suggest that DRA models have only recently emerged as an alternative due to improvements in GIS and spatial modeling software that can more easily calculate spatial-referenced variables but they will become more widely used and will be valuable for public policy evaluations.

Darmofal and Strickler offer up an interesting observation about the lack of interdisciplinary collaboration in their chapter *Bringing Together Spatial Demography and Political Science: Reexamining The Big Sort* (Chap. 8): "while individuals often migrate, ideas rarely do between disciplines." They argue that mid-level theorizing in both political science and demography could benefit from considering how both disciplines can come together over the shared topic of migration and they illustrate this with a re-examination of Bishop and Cushing's (2008) book, *The Big Sort*. Darmofal and Strickler argue that Bishop's analysis and interpretation is limited by a lack of attention to migration scholarship; specifically assumptions regarding the drivers of migration, inattention to the mechanism that fuels political change, a focus on internal migration to the exclusion of international migration, and an inappropriate level of analysis for studying migration. To illustrate the potential gains from cross-pollination between disciplines Darmofal and Strickler suggest that while political geography is central to *The Big Sort* paradoxically spatial concerns—spatial patterning and spatial dependence—play only a minor role. More examples are needed to promote the incorporation of spatial theory across the social sciences.These two chapters, that make up *Part II,* are a bridge to *Part III* which focuses more explicitly on research practice in an attempt to improve our knowledge base in specific areas through the application of spatial analytic tools and methods.

In *Part III Middle-Range Theory in Practice*, we have included eight chapters (Chaps. 9, 10, 11, 12, 13, 14, 15, and 16). Many of the chapters are paired or grouped based on shared subject matter reflecting contributions to the collection from geographers (Chap. 9), rural sociologists (Chaps. 10, 11, 12 and 13), health researchers (Chap. 14), and demographers (Chaps. 15 and 16). Across the collection of chapters we see the infusion of theory and spatial concepts, the innovative use of geospatial data, and the use of different spatial analytical tools, ranging from exploratory spatial data analysis (ESDA) through to advanced spatial econometric models.

Shin and Agnew provide a theoretically grounded and clear empirical analysis that serves to promote the adoption of spatial theory in their chapter, titled *Demography and Democracy: Exploring the Linkage between Age and Voter Turnout in Italy with Geospatial Analysis* (Chap. 9). This chapter explores this linkage between age and abstentionism in Italy (1946–2013), a country with historically high turnout rates and an increasingly aged population. The framing of the substantive questions, the introduction of the data and methods used—ESDA and spatial econometrics—and the analytical strategy including domain specific models (national, northern and southern Italy) are all clearly specified. Shaw and Agnew argue that appreciating and understanding the linkage between age and voting requires theoretical and methodological approaches that are sensitive to global trends, sub-national patterns, and local idiosyncrasies. In closing they note that while neither the local nor the national are privileged in mid-range approaches both are recognized as necessarily complementary.

Howell and Porter's *Decomposing County Population Growth in the United States: Spatial Patterns, New Geographies, and New Methods* (Chap. 10) is an introduction to demographers of a new theoretically meaningful geography for examining the often studied dynamics associated with rural-to-urban (and vice versa) population redistribution. The author's make the point that understanding the dynamics of any demographic of social process that occurs in space should always be done with a special attention to the unit of spatial analysis in which the process is captured or measured. The research builds on both past research at, and criticisms of, the level of analysis of that research. As a contribution to the state of understanding the dynamics of population redistribution, Howell and Porter created a meso-level place/non-place geography in which population counts can be redistributed and aggregated to the phenomenologically meaningful characterizations of being “in-town” or “in the city” versus “out in the country”. This delineation makes a theoretical contribution in that it measures space in a way that is meaningful to the populations that are being measured and in a way that is more telling of redistribution dynamics when compared to past arbitrary boundary approaches (i.e. county aggregates, census tract’s as neighborhoods, etc.). Finally, this chapter introduces the use of a bivariate form of the LISA clustering statistic to demographic analysis as a way of capturing the dynamic mobility of population flows over a given period of time.

In *Socio-spatial Holes in the Advocacy Umbrella: The Spatial Diffusion of Risk and Network Response among Environmental Organizations in the Marcellus Hydro-fracturing Region* (Chap. 11) **Irwin and Pischke** use spatial gravity models to examine the effect of the spatial distribution of hydro-fracturing activity in Pennsylvania on the formation of networks of interaction among environmental advocacy organizations. The scope, structure and density of these networks constitute important dimensions of mobilization in social movements, but are seldom analyzed along spatial dimensions. As the chapter title implies Irwin and Pischke find holes in the structure of this network that leaves specific communities underserved and more at risk for environmental impacts. The authors identify potential areas for collaboration between social movement and socio-spatial approaches, as well as a discussion of implications for mobilization and advocacy from a spatial network perspective.

In *American Civic Community Over Space and Time* (Chap. 12) **Tolbert, Blanchard, Mencken, and Li** examine county-level civic community—the social and economic structures and institutions that buffer communities from external, usually global, forces—with an explicit focus on changing spatial patterns over time (post-1980). To date cross-sectional research has been unable to address whether social capital is in actual decline (Putnam 2000), and the aspatial nature of the research has ignored the possibility of local variation in civic community. Tolbert and colleagues test for declining levels of civic community with a theoretically informed analysis while controlling for spatial unevenness in the distribution of civic community, based on fixed effect model for panel data that integrates a spatial lag term. They find considerable variation across time and space that cast doubt on generalizations about secular decline in American social capital and civic institutions. As noted by other authors in this collection, this type of research reveals how sub-national analyses can help revitalize research on spatial inequality (Lobao et al. 2007).

The link between the theoretical framing and analytical sophistication is maintained by **Yang, Shoff, and Noah** in *Revisiting the Rural Paradox in US*

Counties with a Spatial Durbin Modeling (Chap. 13). The rural urban paradox that is examined refers to lower than expected standardized mortality rates found in rural counties, despite the poor socioeconomic conditions and health infrastructure. Yang and colleagues identify the theoretical and methodological shortcomings of previous research on the rural paradox—including naïve measurement of rural areas and the concept of rurality—and presents a revised theoretical framework that they test using spatial Durbin models (Elhorst 2010; LeSage and Pace (2009). Spatial Durbin models explicitly account for the exogenous relationships between the mortality of a county and the characteristics of the neighboring counties as well as the endogenous relationships between the mortality and explanatory covariates within a county. Analysis in this chapter included an examination of higher-order spatial lag structures. This chapter is a clear application of the spatial Durbin model and it illustrates the value of utilizing spatial structure to explain, in this instance, the variation of mortality across space.

Chapter 14 continues the focus on health outcomes. In *Race, Place, and Space: Ecosocial Theory and Spatiotemporal Patterns of Pregnancy Outcomes* (Chap. 14) **Kramer** uses black-white racial differences in risk for a pregnancy outcome—low-birth weight-preterm birth—as an example for spatializing Krieger's ecosocial theory (2011) and Geronimus' life-course 'weathering' hypothesis (1996). The empirical analysis uses pregnancy outcome data from Georgia 1994–2007 and explicitly addresses three spatial themes: (i) the question of 'how local is local' is asked by considering women's neighborhoods as defined at multiple scales of census geography as well as with egocentric neighborhoods; (ii) the accumulation of socio-spatial 'exposures' across the life course is considered through the lens of the weathering hypothesis, and (iii) the independent contribution of neighborhood trajectories and temporal dynamics on pregnancy outcomes. The three themes are not the entirety of spatializing social epidemiologic theory but, as Kramer notes, they illustrate the potential to work beyond the traditional static, cross-sectional, and arbitrarily bounded health geographies which dominates the extant literature and in turn can help advance our understanding of place-health relationships.

In Chaps. 15 and 16 we turn to back to international research and two chapters that discuss forms of data rarely used by social scientists. **Chen** is interested in *Using Nighttime lights Data as a Proxy in Social Scientific Research* (Chap. 15). The possibility of constructing proxy measures for variables, relating to migration, population growth, and poverty hold great potential for demographic researchers working in countries with low-quality statistical systems, population census, or surveys. In this chapter satellite-based nighttime lights and existing statistical methods are used to generate a proxy for economic statistics focusing on urbanization and poverty. The chapter concludes that both the proposed methodology and nighttime lights data holds great potential for social scientific research where data availability and quality of data at smaller scales have proven a hindrance in past research. The approach used by Chen not only provide mathematical calculations of optimal weights on proxy measures, but also opens it to formal reliability testing and sensitivity analysis. Moreover, the methods can be applied to test many other types of geocoded data, including a wide range of remote sensing information that has not been fully utilized by social scientists.

As a spatial scientist and demographer, **Parker** is interested in temporal and spatial dynamics; specifically in theoretical models and analytical approaches that incorporate both space and time. In his chapter *Human Migration and Spatial Synchrony: Spatial Patterns in Temporal Trends* (Chap. 16), Parker introduces spatial synchrony. Spatial synchrony is an approach borrowed from population ecology that can be used to analyze spatial correlations in time series data. He draws on migration data on a highly mobile ethnic group (the Karen) in Southeast Asia in order to show how theory and empirical data can be tied together. The second half of the chapter discusses emerging issues related to spatial demography, namely: issues of scale, new forms of data and how to deal with them, and research ethics.

Part III has included a diverse set of chapters but all have included clear examples of spatial thinking, the adoption and operationalization of spatial concepts, and use of spatial analytical methods used to explore or test middle-range theoretical frameworks. In general, the collection of chapters include applications in a number of different disciplinary areas (and fields within those disciplines) with the resulting theme being a focus on the identification of proper spatial units, methods, and theoretical explanations. In particular, we are interested in the final point here as it is clear that understanding social phenomena from a general spatial standpoint is not the most effective manner from which to view these processes. Instead, the ability to understand social processes in conjunction with the development/application of a focused theoretical framework that is grounded in the spatial aspects of topic of inquiry has proved to contribute insight beyond what would be gleaned from non-spatial approaches. In some cases, it is the process itself that is spatial and in others the spatiality of the process is less important than the context in which the process occurs. These chapters have given us a place from which to being to understand the difference and connections between such concepts as each has individually developed and applied middle range theory within a spatially-centered framework.

The final chapters (*Part IV*) of this edited collection include a discussion of *Instruction in Spatial Demography* (Chap. 17) by **Matthews** and the concluding chapter (Chap. 18) by **Howell, Porter and Matthews**. As Matthews notes, many important social and demographic questions deserve to be framed and studied using spatial approaches and this will become even more evident as changes in the volume, source, and form of available demographic data – much of it geocoded – further changes the data landscape and thus the methods demographic researchers need to use. Changes in the data demographers collect, how they collect data, how they link data, and how they analyze data suggest the need to train next-generation population scientists in spatial thinking, concepts, and methods of analysis. The challenge is that the training many social scientists and demographers receive in fundamental spatial concepts, geospatial data, and analytical methods is often limited, patchwork, or nonexistent. Instructional resources (courses, textbooks, software and other resources), few of which focus explicitly on demographic research do exist of course but new directions and strategies may need to be developed to both enhance instruction and raise the visibility of spatial demography. Finally, the concluding chapter revisits the primary focus of this edited volume; advancing thinking in spatial demography through the enhancement of our understanding of middle range theory. In doing so, we summarize contributions made by each of the contributing chapters, we emphasize key themes arising from this endeavour, we discuss continued gaps in our understanding, and we articulate some future directions for the field of spatial demography. We understand that this is an incomplete statement on the “state of the art” but we hope to be one of many contribution to help push our understanding of spatial methods in the area of spatial demography (and the greater social sciences) forward.

References

Bishop, B., & Cushing, R. G. (2008). *The big sort: Why the clustering of like-minded Americans is tearing us apart*. Boston: Mariner Books.

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis, 5* (1), 9–28.

Geronimus, A. T. (1996). Black/white differences in the relationship of maternal age to birthweight: A population-based test of the weathering hypothesis. *Social Science and Medicine, 42*(4), 589–597.

Krieger, N. (2011). *Epidemiology and the people’s health: Theory and context*. New York: Oxford University Press.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics* (Vol. 196). Boca Raton: Chapman and Hall/CRC.

Lobao, L. M., Hooks, G., & Tickamyer, A. R. (Eds.). (2007). *The sociology of spatial inequality*. Albany: State University of New York Press.

Matthews, S. A., Janelle, D. G., & Goodchild, M. F. (2012). *Future directions in spatial demography specialist meeting: Final report*. http://ncgia.ucsb.edu/projects/spatial-demography/docs/Future-Directions-in-Spatial-Demography-Report.pdf

Merton, R. K. (1968). *Social theory and social structure*. New York: Free Press.

Porter, J. R., & Howell, F. M. (2012). *Geographical sociology: Theoretical foundations and methodological applications in the sociology of location* (GeoJournal library, Vol. 105). Dordrecht/New York: Springer.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Weeks, J. R. (2004). The role of spatial analysis in demograohic research. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science: Examples in best practice* (pp. 381–399). New York: Oxford University Press.

Weeks, J. R. (2011). *Population: An introduction to concepts and issues* (Eleventhth ed.). Belmont: Wadsworth Cengage Learning.

Weeks, J. R., Hill, A. G., & Stoler, J. (Eds.). (2013). *Spatial inequalities: Health, poverty, and place in Accra, Ghana* (GeoJournal library, Vol. 110). Dordrecht/New York: Springer.

Part I
Theory, Concepts, and Measures

Chapter 2
Challenges of Spatial Thinking

John R. Logan

2.1 Introduction

There has been a steady growth of interest in a range of concepts and techniques in sociology that can be described as spatial. Much of this builds on a large body of work by geographers, and this review will offer some links to that literature. What is distinctive to sociology (and specifically to demography) is the application of spatial data, measures, and models to a wider range of substantive questions with roots in other intellectual traditions. Sociologists are less interested in spatial patterns in themselves, and more interested in how they translate into social relations.

Writing from the perspective of an urban sociologist, I am particularly attuned to the relevance of *place* to social life. Everything happens somewhere, which means that all action is embedded in place and may be affected by its placement. Abbott (1997, pp. 1152) tells us that this is a specifically Chicago School insight, "that one cannot understand social life without understanding the arrangements of particular social actors in particular social times and places... Social facts are *located*." I believe this insight is not unique to the Chicago School. [First Editor's Note: Indeed, the Chicago School's social ecology can be directly traced to the work of rural sociologist Charles J. Galpin at the University of Wisconsin (Porter and Howell 2012).] Much of my own research in the last three decades is centered on questions of inequalities between places (Logan 1978). In the urban political economy tradition every place is socially constructed with a history and a future; where people are placed affects their fortunes and adds structure to their lives; place-based interests are at the heart of much collective and political action (Logan and Molotch 1987). Nevertheless for the purpose of this essay, the key concept is

J.R. Logan (✉)
Department of Sociology, Brown University, Providence, RI 02912, USA
e-mail: john_logan@brown.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_2

not place but space. And by space, I mean specifically location. Spatial thinking is about where things are or where they happen, and it is especially about *where they are in relation to others*. There is an implicit spatial reference in almost all studies of places (not only in urban settings, see Lichter and Brown 2011). What is distinctive about social science in the last decade is that space is being introduced more explicitly and more systematically.

My purpose in this review essay is to examine how key spatial tools and concepts are being used in sociology. I suggest ways in which greater thoughtfulness in their use can yield better results on the initial questions that motivate researchers to use spatial methods and can also lead into more general issues about spatiality.

In a review essay on spatial demography, Voss (2007) argues that traditional demography through the mid-twentieth century was “spatial” in the sense that it was the study of ecological units like cities and counties. A shift of focus to individuals and individual-level processes (associated with increased availability of data at this level) pulled demography away from its spatial origins until the advent of multi-level modeling gave us methods to distinguish between processes at the individual and aggregate levels. This use of a spatial vocabulary seems to suggest that counties are spatial but people are not. In a more careful formulation Entwisle (2007) treats both people and places as spatial, but calls for an approach that gives more agency to people – emphasizing that people make choices about where to live, that they move and that their movements can collectively result in changes in place characteristics and restructure their social networks. She suggests that the popularity of multi-level models reinforces a “top-down” understanding of the relation between people and places, because they focus attention on dependent variables at the individual level.

Yet (like Voss) Entwisle’s point of reference is place and the literature on place effects, and she uses the term “spatial” mainly to refer to places as local “social and spatial” contexts. Again, I consider place to be a fundamental concept for spatial social science. Counties, villages and other areal units have p*lace attributes* that we certainly want to know about. If “where” is a certain residential district, we want to know whether that district is a blue-stocking neighborhood or a ghetto. But for the purpose of this review I call attention to the other attributes that we want to know from a more *explicitly spatial* perspective (e.g., where it is in relation to other places, is it near the center city or out in the suburbs, is it close to a transit line, how long does it take to get to the daycare center, what else is in the vicinity). Concepts that are critical to spatial thinking – distance, proximity, exposure, and access – are all rooted in relative locations. Questions of location are equally spatial regardless of whether the unit of analysis is a person, a firm, or a city, and studies of place that do not deal with location are less so.

One more observation will help to define what I mean by “location.” By now the widespread use of satellite-based geographic positioning systems (GPS) has made us very aware of location as a set of geographic coordinates. And indeed coordinate systems have always been crucial to systematic mapping. GIS maps that make it possible to visualize spatial patterns and to make the measurements required for spatial analysis absolutely rely on measuring longitude and latitude.

Sometimes "location" refers to these points or to locations that can be represented by them – the location of a school or worksite, a crime incident, a riot, a case of measles. Perhaps more often it refers to a larger territory. We use terms like neighborhood or zone to identify a location that is not a single point or address on a map. Such terms seem natural, they are convenient, and they are necessary to spatial thinking. But they introduce two ambiguities. First, what is the geographic scale of the territory? Political studies often deal with world regions or nation states. River basins and valleys define territories for environmental research. Metropolitan labor markets, cities, and more local areas within cities are important to urban analysis. Second, are these territories bounded? Routinely social scientists deal with unmarked boundaries. It can be unclear whether the notion of a boundary is a social construction in the mind of the scientist or a concept grounded in local usage. Does it make more sense to think of boundaries as sharp edges or extended zones of transition? Administrative units that social scientists use regularly in their research, like census tracts, have established boundaries but they may not have a social meaning. Their arbitrary character has led some geographers to replace them with continuous surfaces that they believe may better represent the underlying spatial distribution of population characteristics. Even when formal political boundaries have been established, so that there is a clear line of demarcation between one territory and another, one can question their impact or permeability.

When our object of study cannot be located as a point but must instead be thought of as a place, spatial analysis requires that we confront questions about what constitutes the place. Here I treat the distinction between space and place with a different emphasis than did Gieryn (2000, p. 465), who dismissed space as "what place becomes when the unique gathering of things, meanings and values are sucked out." "In particular," he said, "place should not be confused with the use of geographic or cartographic metaphors (boundaries, territories) that define conceptual and analytical spaces." In fact places are not only geographically located and material, as Gieryn points out, but they are also spatial and their spatiality gives rise to fruitful questions.

These remarks lead to a demarcation of the scholarship that I will review here. Spatial thinking is the consideration of the relative locations of social phenomena, the causes of the locational pattern, and its consequences. It encompasses phenomena whose locations can be thought of as discrete points as well as larger territories, and in the latter case, it requires that we consider questions that are posed as strictly geographical, like whether and where territories are bounded. In fact, like most questions of method and measurement, the underlying issues are not technical but substantive. This is why as often as possible I use the term spatial thinking rather than spatial analysis. Although some leading geographers (Goodchild 2004) have sought to build from Geographical Information Systems toward GIScience as a distinct discipline, most scholars' interest in space is how to incorporate it usefully into their own research agenda. We can profit as much from seeing how others are thinking about space as we can from the advanced tools that are being made available from GIS and spatial statistics.

2.2 Mapping

The most powerful spatial tool is the simplest – creation of a map that allows visualization of a spatial pattern. There has been an explosion of social science mapping in the last decade, advanced especially by the availability of user-friendly software to make original maps and of systems to display maps with web browsers. Maps, like photographs, can provide many layers of information, much of it implicit. I believe their power comes from the combination of their ability to offer an objective representation and their capacity to call on people's imagination. All aspects of semiotics that are typically applied to other aspects of language and communication are relevant to maps (MacEachen 2004), including how people process visual stimuli, how perception and attention are organized, and the mental categories through which people interpret what they see.

When used in an exploratory way by an analyst who is looking for possible spatial patterns, the map's utility is dependent on the analyst's insightfulness. The analyst creates meaning by making associations between the patterns explicitly shown on the map and other extraneous information. Both expected and unexpected observations lead toward new questions. Simple inspection of maps can be facilitated by the use of techniques developed for Exploratory Spatial Data Analysis (ESDA) that draws attention is given to spatial clusters and outliers, and offers methods to "smooth out" random spatial variation so that non-random patterns will stand out more clearly (Anselin et al. 2004). In the sociological literature one of the most common of these tools is the mapping of values of local Moran's i, a measure of clusters of high or low values on a single variable. Adoption of this method was facilitated by the early publication of studies that relied heavily on it (several of which are cited below).

A published map can also have a specific intention, a point that the maker wishes to convey. An effective map, like a well written paragraph, often has a rhetorical character, and the content of a map is not always "objective" information. Cognitive mapping and community-based mapping are important ways of assembling people's imagination about space, which may not correspond to what an outsider could directly observe. A recent exhibit on "Geographic Knowledge in Greco-Roman Antiquity" (New York University 2013) includes very revealing maps of how the most scientifically minded scholars of that era made sense of quadrants of the globe that they could not directly observe.

Even "objective" maps with precise coordinates have a subjective (that is, a purposeful) element in their construction. Geographers have long been aware that there is a discipline to making clear and easily interpreted maps, as well as pitfalls that can lead to "lying with maps" (Monmonier 2005). Yet good maps, like good writing, do not lie. Rather they selectively direct one's attention and train of thought.

Thematic maps have two functions in communicating research results. One is to reinforce findings on the extent of variation in a place characteristic. Seeing the values of a single variable arrayed on a map provides much information in a succinct form. The second function is to demonstrate that the variation has a spatial pattern, offering cues about where values are higher or lower that the viewer can draw on and that the researcher can reinforce in the text as a first step toward reaching conclusions.

Providing maps for these purposes and without any further systematic spatial analysis is now common and is probably the form in which explicit spatial information is most frequently included in social science writing. An excellent example is a study of where people of different race and ethnicity live and work in the Los Angeles metropolitan region (Ellis et al. 2004). Its conclusion – that groups are typically more segregated in their neighborhoods than in the location of their work – is mainly supported by analysis of segregation indices that use no information about where the concentrations of non-Hispanic whites or Salvadorans are, but only that they are more concentrated in some census tracts than in others. But like many studies dating back to when sociologists had to draw maps by hand, the conclusion is reinforced and elaborated with GIS maps.

These maps show specifically where group members are over-represented or under-represented. For example, two side-by-side maps (Fig. 2.1) show the distribution of native-born whites by where they live and where they work. It is evident that there is a large core area of L.A. where these people are very greatly under-represented *as residents*, labeled by the authors as East L.A. and South Central. These are place names that convey considerable meaning to people with even a vague knowledge of the region. There is an even larger territory in the periphery of the region where they are over-represented. Here the only label that is clearly in such a zone is Malibu, a beach town that also has a clear reputation in American culture. The map of workplaces presents a strong contrast. Almost the entire region (but not East L.A. and South Central) is shown as neither over- nor under-represented, suggesting that whites work almost everywhere in similar proportions. Based on these maps, most readers would have little difficulty reaching the conclusion that whites are more segregated at home than at work, a finding that the segregation indices confirm.

One point that I want to make here is how much information is contained in maps like these, particularly as one map is compared to another, place labels are used to stimulate particular associations, and readers are invited to make sense of the "raw" data for themselves. Of course, another point is that the maps are not the original raw data, but rather they are representations that the authors have thoughtfully constructed to convey their own conclusions about the patterns.

2.3 Distance

Distance – the location of something in relation to something else – is often the key message of a map. It is also the core concept in many sorts of spatial research, so central that it is incorporated in the First Law of Geography, which holds that "everything is related, but near things are more related than distant things" (Tobler

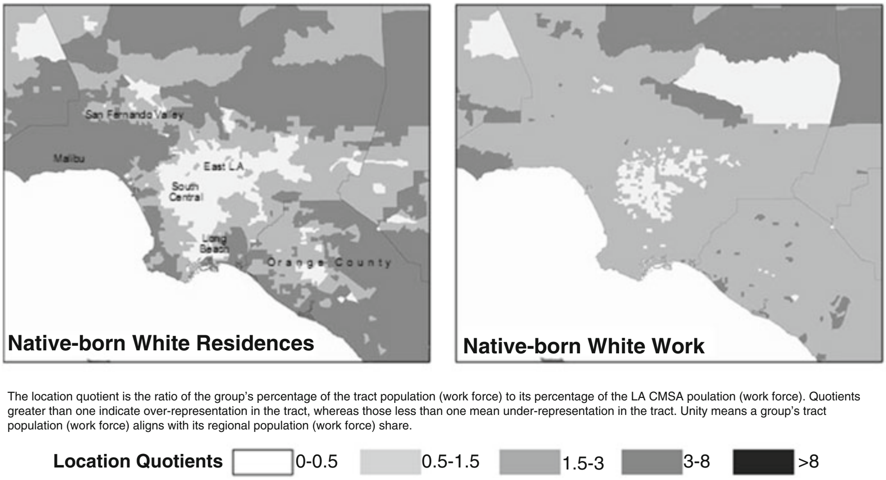

Fig. 2.1 Distribution of native non-Hispanic whites in Los Angeles, 1990, by place of residence and place of work (Source: Prepared by Richard Wright, Dartmouth University)

1970, p. 236; for a recent evaluation see Sui 2004). In the following sections I review research that treats distance as an indicator of access to other people or resources or exposure to harm, as a feature of social networks, and as a basis for identifying and interpreting clusters of related things (see Logan et al. [2010] for more details about the methods used in such research).

2.3.1 *Distance as Access*

Many studies construct some version of what geographers call "egocentric" neighborhoods (Chaix 2009) or local environments, often based on measuring every person's proximity to some place characteristic or to every other person in the system (typically, but not always, based on their place of residence). The notion that every person lives at the center of their neighborhood is reinforced by studies of people's perceptions (Coulton et al. 2001), but typically researchers rely on documentary sources to simulate an egocentric neighborhood. For example Frank et al. (2004) defined a 1-km circle around a person's home to study the effect of the built environment (street networks, land use mix, population density) on travel behavior. It is rare for researchers to have precise locational information (except through confidential databases). Most often, therefore, researchers assign people to a census tract or ZIP code, and pretend that every resident lives at its centroid. An alternative (known as kriging) is to simulate the actual distribution of people (and categories of people) at every point in space. In the latter case analysts proceed *as if* they had exact locational information for every person (e.g., Lee et al. 2008).

Distance naturally arises in transportation studies, which consider how the cost of distance can be managed and how it affects people's behavior. Baum-Snow (2007) asked whether changes in transportation infrastructure – new highways leading from the city center to the suburbs – affected where people live in metropolitan areas. He found that in both 1950 and 1990 there was a strong negative association between population density in census tracts and their distance from an interstate highway, controlling for distance from the central business district. More telling, population growth of central cities during the period (using constant city boundaries) was significantly lower in metropolitan areas with more or increasing numbers of highways radiating out of the city center. Baum-Snow and Kahn (2000) conducted a similar analysis of mass transit access and use. They showed that neighborhoods closer to new transit access points (looking only within a 2 km radius) experienced a significant shift toward commuting via mass transit. But in the metropolitan areas that they studied, expansion of transit lines during the 1980s occurred mainly in the suburbs and improvements were experienced especially in census tracts with more homeowners and college-educated residents, fewer African Americans, and fewer young adults. This finding points toward a concern with spatial inequality (unequal access or exposure by different population groups) that is often the motivation for spatial analysis (Lobao et al. 2008).

A number of social scientists have looked at proximity to employment concentrations in terms of spatial inequality. A familiar concept in the deindustrialization literature is spatial mismatch between people and employment (e.g., Kain 1992). The core idea is that increasingly the metropolitan job base, especially for lower skilled jobs, has shifted to the suburbs, while population groups in need of such jobs (especially young minorities with lower education) are disproportionately found in central cities. Distance from jobs becomes an impediment to being employed. An early effort to test this notion (Cohn and Fossett 1996) evaluated the spatial location of employment in Boston and Houston, particularly in entry-level blue collar positions. The analysis required construction of a measure of employment opportunities that is specific to people living in a given census tract. In both cities, in fact, they found that jobs were highly centralized, as was the black population. They then asked whether the proximity of census tracts to employment opportunities was related to their racial composition, and reported that in fact blacks were more highly represented in tracts with greater proximity to jobs.

Allard and Danziger (2003) asked whether welfare recipients who live in closer proximity to employment opportunities are more likely to find a job and less likely to remain on welfare than those who live in more job-poor locations. They found that white welfare recipients lived in neighborhoods with better access to jobs than did black recipients, largely due to their more suburban residential pattern. Regardless of race, those with better job access were more likely to get a job and to exit welfare.

Another example of interpreting proximity in terms of access (where being closer is regarded as a positive attribute) is a study of the location of banking outlets in Milwaukee. Squires and O'Connor (1998) documented the avoidance of poor and minority neighborhoods by regular bank branches and their replacement by check-cashing firms and other fringe banking outlets, which they interpreted as inferior service.

2.3.2 *Distance as Exposure or Risk*

Proximity can also be treated as exposure to potential harm. This is illustrated in studies of environmental justice, many of which ask whether minorities – by virtue of their location – are disproportionately exposed to pollution. Pastor et al. (2004) studied the spatial distribution of sites in California that were known to release toxic emissions. Emissions sites were categorized as facilities reporting any type of toxic air release, facilities reporting emissions of "persistent bio-accumulative toxins" (PBT), and sites reporting releases of an EPA priority category of toxics known as 33/50 chemicals. These sites were geocoded, and GIS methods were used to create circular zones (distance buffers) of ½ mile, 1 mile, and 2½ miles around each site. Then census tracts were categorized according to their zone of proximity to these exposures. Note that census tracts were accepted as proxies of neighborhoods, and the question was how close (in three distance intervals) various types of neighborhoods were to a risky location. The key finding is that neighborhoods with high proportions of Latino residents were most exposed.

A more sophisticated approach is to use a distance decay model, where "exposure" to a site is assumed to be proportional to one's distance from it. Downey (2006) used this method to examine whether minority and lower income groups are disproportionately burdened by environmental hazards in Detroit. He began with a map showing the geocoded location of industrial facilities identified in the federal government's 2000 Toxics Release Inventory. He overlaid this map with a census tract map, and calculated the distance from every toxic facility to each of many small grids within each tract. He then calculated the total hazard exposure for each grid, taking into account these distances and also the volume of toxic emissions from each facility, and aggregated the grid cells to calculate a total tract exposure. There are two hurdles for this analysis. The first is that Downey did not know what variation there was in population composition of the many grid cells within each tract. He chose to presume that they were all the same. The second hurdle was to assess how distance should be related to exposure – should exposure decline linearly with distance, or should nearby facilities be counted even more heavily than more distant ones, and is there some distance beyond which there is no exposure? Because there is no obvious solution, Downey chose six different distance-decay functions and tested all of them. He used multiple regression analysis to determine that the percent of black residents in a tract is significantly related to toxic exposure, but only at distances of 1.5–2.5 miles. Black census tracts tended to be near but not directly adjacent to toxic facilities. Without a stronger theory about the expected distance band, the significance level of this finding is in doubt – if one tests several cutoff points, there is a probability that at least one of them will appear to be significant even if the distribution is random.

In a more recent study Crowder and Downey (2010) summed the volume of pollutants emitted within 1.5 miles of similar small grid cells. Using longitudinal data from the Panel Study of Income Dynamics (PSID), they were able to show that blacks and Latinos had greater exposure to pollution at the census tract level than whites or Asians, even after controlling for individuals' socioeconomic characteristics. An innovative step was to ask how residential mobility affects this pattern. They found that pollution was not a significant predictor of moving away from a tract for members of any group, but that black householders (and to a lesser extent Latino households) were more likely than whites to move to destinations with higher pollution. Choice of migration destination, then, reinforced racial disparities in environmental exposure.

Pais and Elliott (2008) used similar methods to investigate the effects of another type of environmental risk: three major hurricanes during the early 1990s. What population shifts in neighborhoods (again operationalized as census tracts) were caused by wind damage? This study relied on sophisticated climatological applications of GIS methods to estimate the maximum wind speeds experienced in every census tract within the study region. The researchers combined these estimates with information about the tract's demographic composition (population size, in-migration, and number of housing units) in 1990 (before the storm) and 2000 (afterwards). Their regression procedure adds a special feature that qualifies it as a "spatial regression." To control for the fact that census tracts near one another tend to have similar characteristics, and also tend to have suffered similar levels of wind damage, they included a spatial error term in their model to correct for spatial autocorrelation. They also specifically investigated several spatial factors. Most interesting, it turned out that there was robust population growth in all the areas hit by these hurricanes, but especially in those areas just outside the zone of greatest damage. There was, in a sense, displacement of resources to nearby, less damaged zones.

Residential segregation can also be conceptualized as a spatial exposure, in this case exposure to other people. Michael White (1983) was among the first sociologists to suggest thinking of segregation in terms of the distances between every two persons in a city. In cities where whites tend to live closer to one another than they do to blacks, and blacks tend to live closer to other blacks than to whites, segregation is higher. Working with aggregated data for areas such as census tracts, these distances cannot be precisely measured, but they can be estimated, and White proposed an index that could summarize the spatial pattern of an entire city in terms of groups' relative proximity to one another. Considerable effort has been devoted recently to ways of applying this approach to the construction of "spatial" segregation measures that are analogous to aspatial measures such as the Index of Dissimilarity, exposure and isolation indices, and the information theory index H (Reardon and O'Sullivan 2004). These measures consider people to have some proximity not only to others in the same block or tract, but also to those who live in nearby areas.

2.3.3 Measuring Distance

Key to all studies of access, exposure, or segregation is the question of how to measure distance. Cohn and Fossett's (1996) study of access to jobs tackled this issue directly. They believed that proximity to jobs involved not just Euclidean distance but also how long it actually takes to get from point a to point b. This in turn depended on the mode of transportation (walking, public transit, auto), which is conditioned by the availability of public transit and automobile ownership. And because whites and blacks in the cities they studied, and in specific neighborhoods in those cities, may have very different transportation options, they recognized that creating a valid measure of "how far a neighborhood is to jobs" is not nearly as simple as measuring distances. Their solution was to develop estimates based on several different scenarios about how transportation access varies across races.

Setting aside the problem of the metric for distance (Euclidean distance, distance along a road network, travel time, travel cost) the usual assumption is that there is some limit beyond which more distant locations are irrelevant (often referred to as a "band width") and also some distance-decay function $f(d_{ij})$ that evaluates how much more the nearer points matter in comparison to less near points within that range. But how much more should nearby locations be counted, and at what point is the distance too great to matter? Researchers have approached these questions in various ways.

The most satisfying approach is to find a substantive basis for a given choice. Allard and Danziger (2003) used information on commuting patterns in the Detroit region – how long people typically travel between home and work – to help calibrate their decay parameter (that is, how much less weight they give to employment in more distant tracts). However in many cases there is no relevant data. An alternative is to present results using several alternative assumptions. White (1983) offered four different calculations based on a linear decline, two different exponential rates of decline, and decline with the square of distance, and these resulted in slightly different rank-ordering of cities by level of segregation. Downey (2006) similarly evaluated various functions for exposure to pollutants, including two alternative distance-decay curves (curvilinear and inverse curve) at three cutoff distances (.5, 1.5, and 2.5 miles). Like White, he found different results using different functions, which emphasizes that these choices can be consequential.

One could imagine methods of deciding which of these results is the "real" finding. In segregation research, for example, suppose one's underlying interest were to use geographic data to infer interpersonal contacts within or across racial/ethnic boundaries. In that case one might interview residents to discover at what geographic range they actually notice or talk to other people in their vicinity. This is how Grannis (2009) determined that a single street segment is the key building block of neighborhoods. In environmental studies, suppose that the underlying motive is to estimate health risks from air pollution. Then possibly studies of health outcomes at varying distances from major polluting sources could provide a basis for a decay parameter. Unfortunately we have little knowledge at this level of specificity. Scholars anticipate generally that proximity to pollution is risky, but there is little basis for arguing that one distance function is superior to another.

An interesting alternative in segregation research is to measure exposure not as a single summary statistic but rather as a curve where values of some exposure (e.g., the percentage of Latinos within a given distance of the average Latino resident) are shown to vary with distance. Typically one would expect that people live in relatively more homogeneous social environments at short distances (such as next door neighbors or people on the same black face) but that exposure to co-ethnics ("isolation") declines at greater distances. This is a widely used technique in geography, and it has been proposed as a method to describe the spatial pattern of segregation. Reardon et al. (2008) calculate "spatial" segregation at various cutoff distances, and refer to the resulting curve as a segregation profile. Regions (cities or metropolitan areas) may vary from one another not only in how high the level of isolation is at short distances (such as the spatial scale that could correspond to a typical block or census tract), but also how quickly it declines with distance. The segregation profile offers much more information on the spatial pattern of segregation than a single summary measure.

An objection could be that it provides too much information, since the full segregation profile is not easily summarized and compared across cities or regions. Perhaps for this reason Lee et al. (2008) emphasize two alternative scales that they argue have distinctive meanings. Estimating a segregation measure based on a distance-decay function that reaches no further than a 500 m radius, they propose, corresponds to the case of a pedestrian neighborhood in which walking the dog or taking children to a playground can be done on foot. A radius of 4000 m, in contrast, is larger than many suburban municipalities, and they treat this as the limit of what residents might consider a neighborhood. Whether these are appropriate scales is an important research question. Possibly the "right" scale will vary across metropolitan regions, depending on the density or development in the region and usual methods of transportation.

In my own recent research I have had the luxury of working with geocoded point data showing where everyone lived in several cities in 1880. This freed me and my collaborators to calculate measures of segregation at multiple spatial scales, each of which likely has different meaning as a social context. The finest scale is a multi-unit building. This is not what most social scientists think of as a neighborhood, but in fact newly arriving minority groups often occupy whole apartment buildings. An extreme case is Bijlmermeer in the southeast corner of Amsterdam. Built in the 1970s for a middle-class market, it risked abandonment when that plan failed. But this coincided with a large wave of Surinamese immigrants, and this whole complex became largely Surinamese and Antillean (Logan 2006). In Chicago in 1880, when the black population was quite small, there were no large neighborhoods and only a handful of single street segments that were predominantly black. But at the building level segregation was nearly complete. Figure 2.2 illustrates a mixed block on Chicago's South Side. Most buildings were all-white. One building was equally divided with two white women who employed two black domestic servants. The remaining buildings were majority or entirely black. On average blacks in Chicago

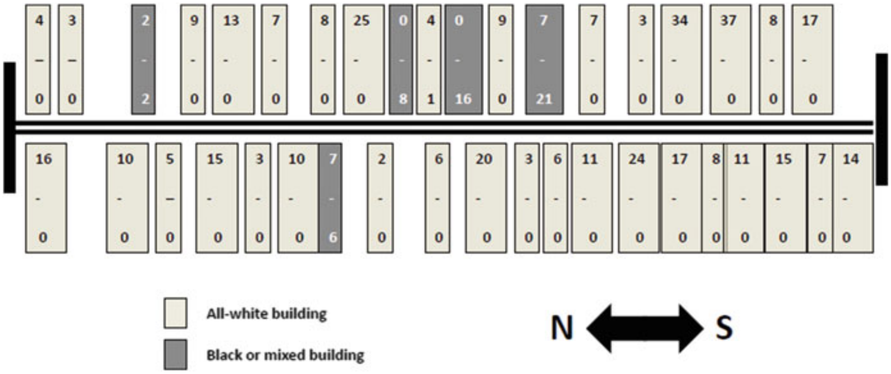

Fig. 2.2 South State Street, Chicago, 1400 block (partial) in 1880, showing number of white residents (*upper numeral*) and black residents (*lower numeral*)

at this time lived in buildings that were 74 % black, but on street segments that were more than three quarters white (Logan et al. 2015).

A fundamental concern with most research that uses distance measures to assess exposure or access is that we take for granted that the relevant measure is distance from one's place of residence. This assumption is convenient, because standard data sources identify place of residence. However some research makes the case for alternative conceptions. For example, Pastor et al. (2002) argue that for children's exposure to environmental pollution, it is at least equally relevant to do measurements in relation to their school. More broadly, Kwan (1999) raises the possibility of an entirely different formulation based on where people actually go on a regular basis – home, work, shopping, and other activities. Using time diaries for Columbus, OH, Kwan seeks to identify the area of the city that is within the "potential path accessibility" of people given the constraints of where they live, where they work, what other routine activities they must perform, and when activities must be performed. In this particular study she demonstrates that women – because their domestic and childcare activities place more constraints on their use of time – have much smaller fields of activity than do men.

2.4 Distance and Social Networks

Proximity also plays a role in social networks, and increasingly analysts of social networks are considering how the spatial arrangement of actors influences the relationships among them. In a variation of Tobler's Law, McPherson et al. (2001, p. 430) observe that "the most basic source of homophily is space: We are more likely to have contact with those who are closer to us in geographic location than those who are distant."

There are alternative views, such as Wellman's (1994) notion of the "liberated" community in which people have flexibility in finding shared interests with others. Ties, he argues, are not established primarily in their residential neighborhood but increasingly through secondary associations not connected with either home or work, and more recently through electronic media (Wellman 2001). Fischer (1995) also observes a high degree of personal choice in social networks formed in urban areas, leading to the emergence of urban subcultures, but he notes that subcultures are sometimes reinforced by spatial clustering. Even ties created through the internet may have a spatial pattern (Hampton and Wellman 2000).

Proximity is certainly not the only factor at work. Britton's (2011) research on neighborhoods in Houston shows that non-Hispanic whites who lived in more integrated neighborhoods actually had *fewer* black friends, except in the case that their black neighbors had high socioeconomic status. But blacks in more integrated neighborhoods had *more* white friends even controlling for their own social class and that of white neighbors. Studies of support networks report that people's closest exchange relationships are not with neighbors but with family members. On the other hand, family members (especially parents or adult children) who live nearby exchange more support and have more frequent contact with one another – the combination of kinship and proximity is especially potent (Logan and Spitze 1994).

Spatial proximity is so closely interlinked with network connections that spatial effects are sometimes interpreted in terms of social interaction. Entwisle et al. (1996) found large differences across Thai villages and high homogeneity within villages in choice of contraceptive methods, higher than could be accounted for in a multilevel model that included individual and place characteristics. They reasoned that "village boundaries largely coincide with the boundaries of social networks" (1996, p. 9) – networks within which women routinely exchanged information about intimate topics. This conclusion was consistent with focus group interviews that probed women's interaction patterns. Liu et al. (2010) noticed that children living very close to a child previously diagnosed with autism were more likely to be given the same diagnosis. They considered several mechanisms that could produce this spatial effect. There could be shared risk from an environmental toxin or contamination of a virus that increases the likelihood of autism. There could be neighborhood selection based on some other personal or family characteristic that induces clustering of risk factors for autism, such as socioeconomic status or parental age. Or there could be a social influence effect resulting from exchange of information between families about symptoms and possible responses to autism. Several findings strengthen the social influence interpretation. In particular, when proximate children lived in different school districts (which would tend to limit parental interaction) the effect of distance to a diagnosed child disappeared. The researchers also took advantage of information about residential mobility. For example, the effect of proximity disappeared if the diagnosed child's family moved away.

Some researchers have sought to assess the relative strength of spatial proximity and social networks when independent measures are available for each concept. Whittington et al. (2009) point out that "social structural spaces" and "geographic agglomeration" are not independent. For example studies of industrial districts suggest that one of their advantages is that proximity fosters social connections among employees of otherwise unrelated firms. In their study of biotechnology firms, Whittington, Owen-Smith, and Powell show that innovation (the number of patents issued) is related to measures of physical proximity (the average distance of the firm to all other firms), centrality in relationships with other firms (based on reported formal contractual agreements), and location in one of the three key regions for this industry (San Francisco Bay Area, Boston, and San Diego). Radil et al. (2010) research on gang-related violence, which explicitly seeks to "spatialize social networks," concludes that both social network connections (rivalries with other gangs) and proximity (treating adjacent territories as their turf) influence violence. However among many similar studies of health risks, Giebultowicz et al. (2011, p. 1387) find that cholera in rural Bangladesh "always clusters in space and seldom within social networks" (as defined by kinship networks).

Wineman et al. (2009) looked at similar questions at the micro-level, examining collaborative innovation by faculty in a professional school, where successful innovation was measured by the number of co-authored articles by each pair of faculty. Being in the same academic department (the indicator of common social network position) had the stronger effect on coauthorship. But proximity (a complex measure of walking distance and line of sight connection between offices) also had a significant impact.

2.5 Spatial Clustering

Another way that distance is incorporated into social research is through the phenomenon of spatial clustering, the pattern of related things being found in proximity to one another that Tobler called attention to. When we refer to clusters, we are typically calling attention to zones in which there is a larger than expected concentration of some characteristic. Clustering is the main focus of Weeks' (2004) review of how spatial analysis can be used in demographic studies. An early study of the spatiality of crime (Sherman et al. 1989) called attention to the existence of crime "hot spots," pointing out that many types of crime are both rare and spatially concentrated at certain locations (based on statistical measures such as Moran's i). Criminologists have been a major contributor to methods of identifying clusters of point-level phenomena, facilitated in part by specialized software for this purpose such as Crimestat (Levine 2010). Health researchers, who often have access to point data on health conditions, have also shown much interest in spatial clusters. One study of dengue infection in a city in central Brazil (Siqueira et al. 2004) used survey data on dengue infection status, medical condition history, and socioeconomic and demographic characteristics, and evaluated whether there were zones of the city with significantly high concentrations of infection, which could then be targeted for public health interventions. Similar techniques were used to identify clusters of childhood leukemia in west-central Lancashire in England during 1954–1992, based on analysis of the proportion of leukemia cases at various distances from each child (Gatrell et al. 1996).

Distance is the starting point for identifying clusters, and all of the questions about what distance means and how to measure it apply to questions about clusters. An additional central issue is whether what appears to be a "hot spot" could be the result of a random process, which requires an understanding of the statistical properties of spatial distributions. For example there was considerable controversy in New York State when detailed information was made public on the location of cancer cases. The raw data suggested that there might be concentrations of risk in sections of Long Island, but experts disagreed on whether these were real or random (Jacquez and Greiling 2003).

2.5.1 Neighborhoods

Whether based on point data or areal units like blocks or census tracts, another application of cluster analysis is to identify "natural areas" (neighborhoods or regions) based on the composition of smaller units. In such work we presume that the spatial cluster is neither random nor ephemeral, but rather represents the existence of meaningful places. And we can ask specifically where these places are, what they are like, and – perhaps most difficult – what are their boundaries.

Like any social category (Lamont and Molnar 2002) we expect neighborhoods to have both symbolic and social boundaries, and sometimes to have political boundaries as well. These are meaningful and often contested. In an extreme example of a divided city, Shlay and Rosen (2010) describe the clash of alternative narratives that support or reject Israel's effort to shift the Green Line between Jewish and Palestinian zones of Jerusalem. In more typical cases, Suttles (1972, p. 4) argues that residents tend to construct simplified images of the city in which differences between neighborhoods, and hence their boundaries, are magnified. Nonetheless researchers have noticed that it is common for borders to be fluid. Hunter (1974), for example, reported that areas of Chicago that he studied had "rolling" boundaries – people might agree on the name of their neighborhood, but those living near its edge tended to perceive it as extending further in that direction. Recent ethnographic studies of cognitive maps show that residents draw on many of the same characteristics to identify the boundaries of their neighborhoods as do social scientists. Lacy (2007) describes this as "boundary work," and focuses on how residents of middle class black areas seek to signify that their neighborhood is distinct from neighborhoods of black working class or poor people. In a Baltimore neighborhood Rich (2009) found that white residents used both the racial composition and class background of specific blocks to mark the limits of their own neighborhood, emphasizing that the more middle-class and white areas were the heart of their neighborhood. But in another locale Campbell et al. (2009) reported that white residents preferred to draw wider boundaries in order to think of their neighborhood as more racially and occupationally diverse.

A simple thematic map of a population characteristic provides a first approximation of where people of different kinds are geographically concentrated. Some of my own research has sought to be more concrete. Alba et al. (1997, p. 892) operationalized an ethnic neighborhood as “a set of contiguous tracts, which must contain at least one tract where a group is represented as 40 % or more of the residents and whose other tracts each have a level of ethnic concentration among residents of at least 35 %.” But when we wished to study ethnic neighborhoods of newer immigrant groups in New York and Los Angeles, such as Chinese or Filipinos, only a handful of census tracts met this criterion, and much of Los Angeles would be defined as a Mexican neighborhood by such criteria. Thematic maps showed visible concentrations of several immigrant minority groups that typically extended across many tracts (Logan et al. 2002). This led us to propose the use of local spatial clustering at the tract level (the same local Moran’s i as used in studies mentioned above to identify spatial clusters of crime or disease) to identify statistically significant clusters, and we treated these larger areas as the groups’ ethnic neighborhoods (see also Logan and Zhang 2004). Grannis (1998) reaches a somewhat similar result by a different route, proposing to identify what he calls t-neighborhoods within areas bounded by major streets and validating this designation with evidence of social homogeneity within these areas.

2.5.2 *Establishing Boundaries*

More recently I have experimented with other spatial clustering techniques using a geocoded data set for Newark, NJ, in 1880 (Logan et al. 2011). This approach frees us from having to accept arbitrary administrative units as the building blocks of ethnic neighborhoods, but we are far from being able to know when a given solution is the right one. In this analysis we chose to use street segments (a single street between two intersections) as the building block for neighborhoods. Figure 2.3 shows the result of a Bayesian model that estimates the probability of a segment being part of a German, Irish, Yankee, or ethnically mixed neighborhood. A key assumption is that segments near one another are more likely to be in the same kind of neighborhood.

There are fundamental conceptual issues that need to be resolved in using such approaches, such as whether areas with diverse populations should be seen as transitional zones between neighborhoods (fuzzy boundaries) or as a distinctive neighborhood type in their own right. Additional information such as the location of local ethnic institutions and services or residents’ social networks is needed to understand the nature of boundaries between areas.

It is convenient when there are strong political boundaries between zones – and when this particular kind of boundary is theoretically relevant. Political boundaries have special significance because they can be very real and concrete in their effects, allocating different resources and rights to people on either side of the line. Lichter et al. (2007) investigated how the drawing of the line itself may depend on

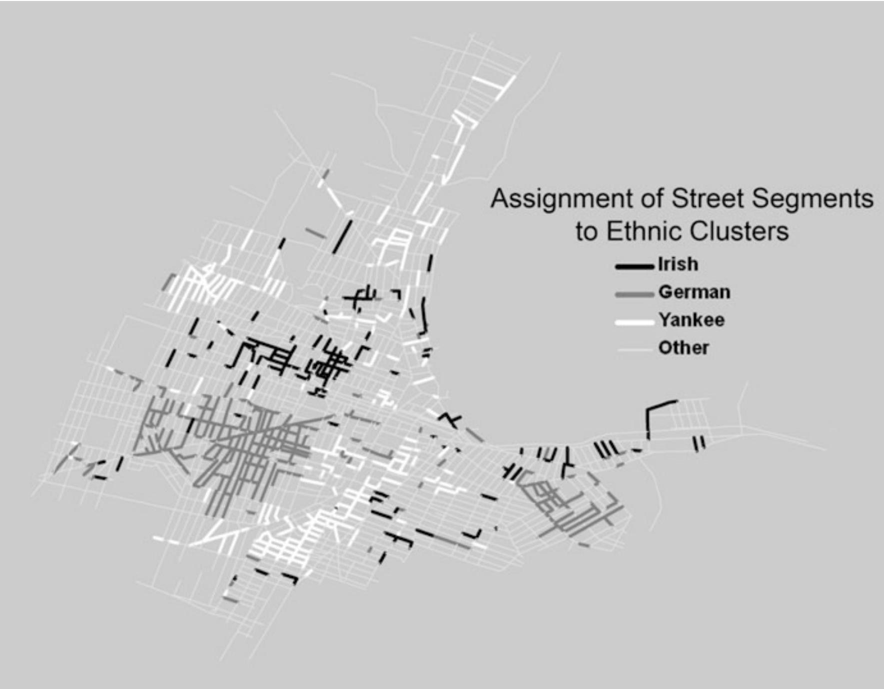

Fig. 2.3 Classification of street segments for Irish, Germans, and Yankees in Newark, 1880

intentions about whom to serve. Their study of annexation in small Southern towns suggests that some areas, based on their racial or class composition, tended to be avoided when town officials extended their borders. Boustan (2007) asked how the border between central cities and suburbs affected housing prices for homes that were otherwise similar (in terms of reported condition, tenure, and density), except that one was on the side of the jurisdiction with a smaller minority population. House prices were higher on this side, regardless of the proportion of minority residents on the block. Boustan estimates that one quarter of this “homogeneity premium” is associated with a preference for children to attend a high school with more white classmates. The latter interpretation is reinforced by research by Saporito and Sohoni (2007) that compared the composition of the school population in particular public schools and in the attendance areas of those schools. They reported that concentrated poverty in local public schools was strongly influenced by decisions of local families to send their children to private schools in the same city, with substantial impact on the schools attended by minority children: “While the typical black or Hispanic child lives in an attendance boundary in which 36 % of the children are poor, the typical black or Hispanic public school student attends a school in which 63 % of the students are poor” (2007, p. 1246).

2.6 Spatial Dependence as an Effect on Neighbors

Clustering is a form of spatial dependence, the tendency for similar things to be near to one another. In the studies mentioned above, spatial dependence was used to identify significant concentrations or to place boundaries around neighborhoods. Another interpretation emphasizes diffusion across boundaries or impact on neighbors. An early example of this interpretation is a study of county-level homicide rates (Baller et al. 2001). It was observed that homicide rates are spatially clustered, particularly high in the American South. The authors argue that this clustering can provide information about the processes that lead to higher or lower homicide. They estimate a multivariate model in which the homicide rate is predicted by several variables believed to be linked to homicide in the criminology literature, such as resource deprivation, and then they analyze the spatial clustering of residuals from this model. Two processes are considered. If the covariance of model residuals with the values of homicide in neighboring counties is high, this is understood as evidence that neighboring areas in fact have an influence even after controlling for the measured predictors. This is termed "spatial lag," and it is dealt with by introducing the homicide rates of adjacent counties directly into the model as an additional predictor. If the covariance of model residuals in an area with model residuals in neighboring areas is high, this is considered to be evidence that clustering is due to the effects of unmeasured predictors. This is termed "spatial error," and it is treated with the regression error term.

Baller et al. added another step to the analysis, asking whether the causal processes are the same in the South and non-South parts of the country (which they term "spatial heterogeneity"). This issue is commonly associated with an analytical technique called geographically weighted regression (Fotheringham et al. 2002). Based on several statistical tests they decided to study the South and non-South as separate "homicide regimes." In the South, they showed that a number of substantive predictors have significant effects but do not account for the residual spatial clustering. They interpreted this evidence of spatial lag in terms of "diffusion," appealing to a number of prior studies that hypothesized that violence can diffuse across space.

Another study by Baller (Baller and Richardson 2002) reported a spatial lag effect in analyses of the geographic clustering of suicides in French departments in the late nineteenth century. They asked whether clustering is due to the fact that the predictors of suicide (such as measures of social integration) are themselves spatially clustered (the position argued by Emil Durkheim at the time) or to the influence of events in one locale on behaviors in adjacent areas (which they describe as the "imitation" hypothesis of Gabriel Tarde). The data seem to support the imitation hypothesis, though the authors emphasize that more direct evidence is needed to show how events in one place actually influence later events in another place – the sort of evidence that is rarely offered in such research. An exceptional case is a study of homicide in Pittsburgh during a period of rapidly increasing youth violence in 1987–1995. Cohen and Tita (1999) take advantage of annual reporting of the location of homicides in the city to examine spatial dependence in homicides over time (i.e., whether homicides at time 2 are spatially associated with homicides at time 1 in adjacent areas). They find evidence of "contagious diffusion" only at the peak of the homicide epidemic, "when high local rates of youth-gang homicides are followed by significant increases in neighboring youth-nongang rates" (p. 491).

2.7 Spatial Dependence in Multi-level Models

Multilevel models, which have become much more widely used in the last decade, treat spatial dependence as a statistical problem. The simplest multilevel question focuses on contextual effects, the impact that characteristics of places or other kinds of contexts have on people or other kinds of actors within them. It has become common, for example, for large scale survey datasets such as the PSID to include geographic identifiers so that researchers can use place characteristics as predictors of individual outcomes. The notion is that something about the context, such as its racial composition or poverty level or density, has a direct impact on people within it. Hierarchical linear modeling (HLM) treats the commonality among people in the same locale as a violation of the assumption of independent observations that is required in ordinary least squares (OLS) regression. HLM estimates standard errors for the coefficients of contextual variables in these models that are corrected for the (spatial) dependence across individuals.

2.7.1 Adding Spatial Lags to Multi-level Models

A more complex situation is when one is interested in contextual effects but also wishes to take into account spatial dependence of individual outcomes across nearby places. Early efforts to conceptualize and deal with this sort of issue are associated with the Project on Human Development in Chicago Neighborhoods (PHDCN). I will focus here on a single seminal publication from that research, where the substantive question is how to explain variations in residents' reports of what the authors call "collective efficacy for children" (Sampson et al. 1999). For simplicity, I will refer to only one indicator of collective efficacy – informal social control – which refers to the expectation that neighborhood residents will intervene in children's misbehavior.

The core hypothesis takes the same form as contextual effects that can be estimated through HLM. It is proposed that there is variation among individuals in their evaluation of informal social control in their neighborhood, based on such factors as their race, socioeconomic status, and length of residence in the area. In addition, there are aspects of the local social structure that lead some neighborhoods to be perceived as having more social control than others. The authors view these contextual characteristics, including residential instability and concentrated disadvantage, as factors that can reinforce or disrupt effective social networks and collective capacity, over and above the effects of similar factors measured at the individual level. This combination of individual and neighborhood effects is well suited for multilevel modeling.

The next theoretical step is to introduce a more explicitly spatial dynamic. The researchers propose to take into account what they call the "embeddedness" of neighborhoods (that is, the relevance of knowing the nature of the larger area within which the neighborhood is located). Figure 2.4 reproduces a map identifying clusters of neighborhoods with higher (high-high) and lower (low-low) child-centered social control, based on the local Moran's *i* statistic. Clearly there is spatial patterning. The authors interpret it in the following way: "If social capital is truly relational, then research that considers neighborhoods as islands unto themselves misses the theoretical point ... [T]he resources in one neighborhood are linked to those in surrounding neighborhoods." Furthermore, "spatial 'flows' for dimensions of social capital are also theoretically compelling because social networks and exchange processes cross the artificial boundaries" of census tracts (1999, p. 637). These influences create positive and negative externalities for adjacent areas. And in fact, this study showed evidence that child social control in a given neighborhood is as much related to social control in surrounding neighborhoods as it is affected by such internal neighborhood characteristics as concentrated disadvantage.

A series of analyses from the PHDCN (see also, for example, Morenoff (2003)) demonstrate the promise of studies that consider both individual-level and contextual effects, as well as what Sampson, Morenoff, and Earls referred to as embeddedness. Chaix et al. (2005), studying mental disorders in Sweden, suggest more complex methods that depend on being able to assess what they call spatially adaptive rings at various distances around individuals, showing much stronger effects of nearby areas than could be found using fixed administrative units. Similar datasets are becoming more widely available (such as LAFANS and AdHealth), and we can anticipate that more researchers will become interested in such spatial effects.

2.7.2 Spatial Effects and Spatial Scale

As in the studies of homicide and suicide discussed above, Sampson et al. (1999) use spatial effects as evidence that behavioral patterns can be transmitted across space to neighboring areas. There is an alternative interpretation, which illustrates how strongly the interpretation of complex model results depends on theory. The authors are careful to note (1999, p. 645) that "spatial dependence also arises as a result of the often-inexact correspondence between the neighborhood boundaries imposed by the census and the ecological patterning of social interactions." In other words, possibly census tracts are the wrong scale of analysis or they create artificial boundaries between areas that should better be understood as in the same

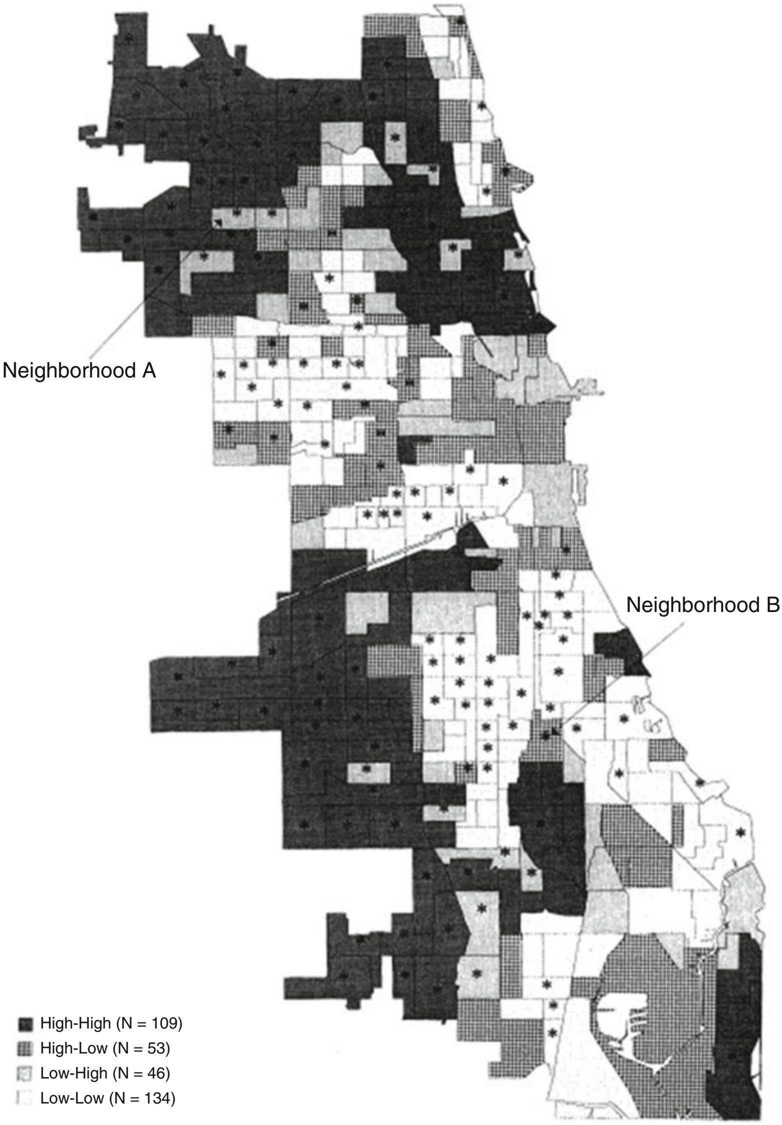

Fig. 2.4 Chicago neighborhoods classified according to spatial clusters of child-centered social control (Source: Sampson et al. 1999)

neighborhood. Scholars familiar with Chicago are aware that there are very large expanses of territory that are occupied by lower income African Americans, or middle class whites, or Latinos. Maps from PHDCN research show that other place characteristics, such as crime rates or informal social control, also extend over large areas, and that these overlap to a large extent with the mapping of race/ethnicity and class. Suppose we thought of spatial dependence as a tool for identifying neighborhood areas, as in Logan et al. (2002)? Then we might expect that Chicago's 865 census tracts could be combined not as 343 "neighborhood clusters" (as in the PHDCN, between 2 and 3 tracts per neighborhood), but more realistically might constitute neighborhoods that are closer in scale to the 77 "community areas" of traditional Chicago ecology (more than 10 tracts per neighborhood). And "spatial lags" in an analysis done at this scale might not be significant. Do spatial lag coefficients mean that behaviors cross into adjacent neighborhoods, or do they indicate that the neighborhood was too narrowly defined? How do we evaluate the appropriate scale for such research?

Studies conducted within metropolitan regions must take a position on the scale of neighborhoods. For example, should neighborhoods be built up from tracts or smaller units within tracts? Hipp (2007) addressed this question by asking at what scale neighborhood characteristics affect residents' experience of crime. He showed that racial/ethnic composition is a more significant predictor at the scale of the census tract than the block, but the effects of average income are greater at the finer scale of the local block.

The usual emphasis on intra-metropolitan differentiation partly reflects the character of the exposures that researchers believe are linked to distance, most of which are not considered to be of importance beyond a certain point (typically not beyond a daily commuting radius). Lobao et al. (2008) criticize this focus with respect to studies of the spatiality of poverty. In their view, there are traditions of scholarship in sociology that particularly privilege two spatial scales. One of these is the local level addressed especially by urban sociology. The other is cross-national and comparative sociology, which especially addresses inequalities between nations. In either tradition the scale tends to be taken for granted, as though it were imposed by the substantive topic being studied. Lobao et al. argue for the importance of the "missing middle" subnational scale (operationalized as differences between regions, states or counties), as does Snyder (2001). A more general point is that sociologists should treat spatial scale as a research question and consider the linkages of processes across scales. We are becoming more aware of how processes of globalization organized at a supra-national scale may be reducing the relative importance of national boundaries and accentuating regional agglomerations and networks of cities (Brenner 2004). The same degree of sensitivity is needed at every spatial scale.

2.8 Looking Forward

Spatial analysis is on a roll. The growing availability of spatially referenced data and the inclusion of geographic identifiers in large scale survey data sets both respond to and reinforce a demand for studies that take place and space more explicitly into account. The basic computing tools are much more easily accessible today than they were in the past, when only specialists could manipulate the programs that make computer maps. Now even general purpose statistical programs include corrections for spatial autocorrelation or clustered sampling, measures of spatial dependence, and procedures for multilevel modeling and spatial regression. The research reviewed is a very partial introduction to the range of spatial questions that are being asked in many areas of social science.

My intention is to promote more sensitivity in use of spatial concepts. Not long ago it was considered to be sophisticated to recognize the ecological fallacy, the point that one cannot infer processes that occur at the level of individuals from information about relationships at the level of places or other social contexts. The distinction between individual and contextual effects is now much better understood. But the difference between place and space with which I introduced this essay may seem obscure, the concept of spatial lag effects in multivariate analysis is new to most social scientists, and it is hard to question how to measure distance when it is only recently that any measure of proximity was in reach. I fully understand Gieryn's impatience with "geographic or cartographic metaphors (boundaries, territories) that define conceptual and analytical spaces" since these may seem to be very formal categories of thought. I hope to have shown that behind these metaphors are significant theoretical and substantive concerns that should guide our use of the new spatial tools at our disposal.

References

Abbott, A. (1997). Of time and space: The contemporary relevance of the Chicago school. *Social Forces, 75*, 1149–1182.

Alba, R. D., Logan, J. R., & Crowder, K. (1997). White neighborhoods and assimilation: The greater New York region, 1980–1990. *Social Forces, 75*, 883–909.

Allard, S. W., & Danziger, S. (2003). Proximity and opportunity: How residence and race affect the employment of welfare recipients. *Housing Policy Debate, 13*, 675–700.

Anselin, L., Kim, Y. W., & Syabri, I. (2004). Web-based analytical tools for the exploration of spatial data. *Journal of Geographical Systems, 6*, 197–218.

Baller, R. D., & Richardson, K. K. (2002). Social integration, imitation, and the geographic patterning of suicide. *American Sociological Review, 67*, 873–888.

Baller, R. D., Anselin, L., Messner, S. F., Deane, G., & Hawkins, D. F. (2001). Structural covariates of U.S. County homicide rates: Incorporating spatial effects. *Criminology, 39*, 561–590.

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics, 122*, 775–805.

Baum-Snow, N., & Kahn, M. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics, 77*, Cambridge. 241–263.

Boustan, L. P. (2007). *Escape from the city? The role of race, income, and local public goods in post-war suburbanization* (NBER Working Paper No. 13311). Cambridge, MA.

Brenner, N. (2004). *New state spaces: Urban governance and the rescaling of statehood*. Oxford/New York: Oxford University Press.

Britton, M. L. (2011). Close together but worlds apart? Residential integration and interethnic friendship in Houston. *City and Community, 10*, 182–204.

Campbell, E., Henly, J. R., Elliott, D. S., & Irwin, K. (2009). Subjective constructions of neighborhood boundaries: Lessons from a qualitative study of four neighborhoods. *Journal of Urban Affairs, 31*, 461–490.

Chaix, B. (2009). Geographic life environments and coronary heart disease: A literature review, theoretical contributions, methodological updates, and a research agenda. *Annual Review of Public Health, 30*, 81–105.

Chaix, B., Merlo, J., Subramanian, S. V., Lynch, J., & Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental, behavioral disorders due to psychoactive substance use in Malmo, Sweden, 2001. *American Journal of Epidemiology, 162*, 171–182.

Cohen, J., & Tita, G. (1999). Diffusion in homicide: Exploring a general method for detecting spatial diffusion processes. *Journal of Quantitative Criminology, 15*, 451–493.

Cohn, S., & Fossett, M. (1996). What spatial mismatch? The proximity of blacks to employment in Boston and Houston. *Social Forces, 75*, 557–572.

Coulton, C. J., Korbin, J., & ChanT, S. M. (2001). Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology, 29*, 371–383.

Crowder, K., & Downey, L. (2010). Inter-neighborhood migration, race, and environmental hazards: Modeling micro-level processes of environmental inequality. *The American Journal of Sociology, 115*, 1110–1149.

Downey, L. (2006). Environmental racial inequality in Detroit. *Social Forces, 85*, 771–796.

Ellis, M., Wright, R., & Park, V. (2004). Work together, live apart? Geographies of racial and ethnic segregation at home and at work. *Annals of the Association of American Geographers, 94*, 620–637.

Entwisle, B. (2007). Putting people into place. *Demography, 44*, 687–703.

Entwisle, B., Rindfuss, R., Guilkey, D. K., Chamratrithirong, A., Curran, S. R., & Sawangdee, Y. (1996). Community and contraceptive choice in rural Thailand: A case study of Nang Rong. *Demography, 33*, 1–11.

Fischer, C. (1995). The subcultural theory of urbanism: A twentieth-year assessment. *American Journal of Sociology, 101*, 543–577.

Fotheringha, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.

Frank, L. D., Engelke, P., & Schmid, T. L. (2004). Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventive Medicine, 27*, 87–96.

Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers, 21*, 256–274.

Giebultowicz, S., Ali, M., Yunus, M., & Emch, M. (2011). A comparison of social and spatial clustering of cholera in Matlab, Bangladesh. *Health and Place, 17*, 490–497.

Gieryn, T. (2000). A space for place in sociology. *Annual Review of Sociology, 26*, 463–496.

Goodchild, M. F. (2004). GIScience, geography, form, and process. *Annals of the American Association of Geographers, 94*, 709–714.

Grannis, R. (1998). The importance of trivial streets: Residential streets and residential segregation. *American Journal of Sociology, 103*, 1530–1564.

Grannis, R. (2009). *From the ground up: How the layered stages of neighbor networks translate geography into neighborhood effects*. Princeton: Princeton University Press.

Hampton, K. N., & Wellman, B. (2000). Examining community in the digital neighborhood: Early results from Canada's wired suburb. In T. Ishida & K. Isbister (Eds.), *Digital cities: Technologies, experiences and future perspectives* (pp. 194–208). Heidelberg: Springer.

Hipp, J. R. (2007). Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review, 72*, 659–680.

Hunter, A. (1974). *Symbolic communities*. Chicago: University of Chicago Press.

Jacquez, G. M., & Greiling, D. A. (2003). Local clustering in breast, lung and colorectal cancer in Long Island, New York. *International Journal of Health Geography, 2*, 3.

Kain, J. R. (1992). The spatial mismatch hypothesis: Three decades later. *Housing Policy Debate, 3*, 371–460.

Kwan, M. P. (1999). Gender and individual access to urban opportunities: A study using space-time measures. *Professional Geographer, 51*, 210–227.

Lacy, K. R. (2007). *Blue-chip black: Race, class, and status in the new black middle class*. Berkeley: University of California Press.

Lamont, M., & Molnar, V. (2002). The study of boundaries in the social sciences. *Annual Review of Sociology, 28*, 167–197.

Lee, B. A., Reardon, S. F., Firebauch, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review, 73*, 766–791.

Levine, N. (2010). *CrimeStat: A spatial statistics program for the analysis of crime incident locations (v 3.3)*. Houston/Washington, DC: Ned Levine & Associates/The National Institute of Justice.

Lichter, D. T., & Brown, D. (2011). Rural America in an urban society: Changing spatial and social boundaries. *Annual Review of Sociology, 37*, 565–592.

Lichter, D. T., Parisi, D., Grice, S. M., & Taquino, M. (2007). Municipal underbounding: Annexation and racial exclusion in small Southern towns. *Rural Sociology, 72*, 47–68.

Liu, K. Y., King, M., & Bearman, P. S. (2010). Social influence and the autism epidemic. *American Journal of Sociology, 115*, 1387–1434.

Lobao, L., Hicks, G., & Tickamyer, A. (2008). Poverty and inequality across space: Sociological reflections on the missing-middle subnational scale. *Cambridge Journal of Regions, Economy and Society, 1*, 89–113.

Logan, J. R. (1978). Growth, politics, and the stratification of places. *American Journal of Sociology, 84*, 404–416.

Logan, J. R. (2006). Variations in immigrant incorporation in the neighborhoods of Amsterdam. *International Journal of Urban and Regional Development, 30*, 485–509.

Logan, J. R., & Molotch, H. L. (1987). *Urban fortunes: The political economy of place*. Berkeley: University of California Press.

Logan, J. R., & Spitze, G. D. (1994). Family neighbors. *American Journal of Sociology, 100*, 453–476.

Logan, J. R., & Zhang, W. (2004). Identifying ethnic neighborhoods with census data. In M. F. Goodchild (Ed.), *Spatially integrated social science* (pp. 113–126). New York: Oxford University Press.

Logan, J. R., Alba, R. D., & Zhang, W. (2002). Immigrant enclaves and ethnic communities in New York and Los Angeles. *American Sociological Review, 67*, 299–322.

Logan, J. R., Zhang, W., & Xu, H. (2010). Applying spatial thinking in social science research. *GeoJournal, 75*, 15–27.

Logan, J. R., Spielman, S., Xu, H., & Klein, P. N. (2011). Identifying and bounding ethnic neighborhoods. *Urban Geography, 32*, 334–359.

Logan, J. R., Zhang, W., & Chunyu, M. (2015). Emergent ghettos: Black neighborhoods in New York and Chicago, 1880–1940. *American Journal of Sociology, 120*(4):1055–1094.

MacEachen, A. M. (2004). *How maps work: Representation, visualization, and design*. New York: Guildford Press.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

Monmonier, M. (2005). Lying with maps. *Statistical Science, 20*, 215–222.

Morenoff, J. D. (2003). Neighborhood mechanisms and the spatial dynamics of birth weight. *American Journal of Sociology, 108*, 976–1017.

New York University. (2013). *Measuring and mapping space: Geographic knowledge in Greco-Roman antiquity*. Institute for the Study of the Ancient World. Viewed at http://isaw.nyu.edu/exhibitions/space on November 6, 2013.

Pais, J. E., & Elliott, J. R. (2008). Places as recovery machines: Vulnerability and neighborhood change after major hurricanes. *Social Forces, 86*(4), 1415–1453.

Pastor, M., Jr., Sadd, J. L., & Morello-Frosch, R. (2002). Who's minding the kids? Pollution, public schools, and environmental justice in Los Angeles. *Social Science Quarterly, 83*, 262–280.

Pastor, M., Sadd, J. L., & Morello-Frosch, R. (2004). Waiting to inhale: The demographics of toxic air release facilities in 21st-century Californian. *Social Science Quarterly, 85*(2), 420–440.

Radil, S. M., Flint, C., & Tita, G. E. (2010). Spatializing social networks: Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. *Annals of the Association of American Geographers, 100*, 307–326.

Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology, 34*, 121–162.

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographical scale of metropolitan racial segregation. *Demography, 45*, 489–514.

Rich, M. A. (2009). 'It depends on how you define integrated': Neighborhood boundaries and racial integration in a Baltimore neighborhood. *Sociological Forum, 24*, 828–853.

Sampson, R., Morenoff, J., & Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review, 64*, 633–660.

Saporito, S., & Sohoni, D. (2007). Mapping educational inequality: Concentrations of poverty among poor and minority students in public schools. *Social Forces, 85*, 1227–1253.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*, 27–56.

Shlay, A., & Rosen, G. (2010). Making place: The shifting green line and the development of Greater Metropolitan Jerusalem. *City and Community, 9*, 358–389.

Siqueira, J. B., Martelli, C. M. T., Maciel, I. J., Oliveira, R. M., Ribeiro, M. G., Amorim, F. P., et al. (2004). Household survey of dengue infection in central Brazil: Spatial point pattern analysis and risk factors assessment. *American Journal of Tropical Medicine and Hygiene, 71*, 646–651.

Snyder, R. (2001). Scaling down: The subnational comparative method. *Studies in Comparative International Development, 36*, 93–110.

Squires, G. D., & O'Connor, S. (1998). Fringe banking in Milwaukee: The rise of check cashing businesses and the emergence of a two-tiered banking system. *Urban Affairs Review, 34*, 126–163.

Sui, D. Z. (2004). Tobler's first law of geography: A big idea for a small world? *Annals of the American Association of Geographers, 94*, 269–277.

Suttles, G. D. (1972). *The social construction of communities*. Chicago: University of Chicago Press.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography, 46*, 234–240.

Voss, P. R. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*, 457–476.

Weeks, J. (2004). The role of spatial analysis in demographic research. Chapter 19. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science*. New York: Oxford.

Wellman, B. (1994). The community question: The intimate networks of East Yorkers. *American Journal of Sociology, 84*, 1201–1231.

Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research, 25*, 227–252.

White, M. J. (1983). The measurement of spatial segregation. *The American Journal of Sociology, 88*, 1008–1018.

Whittington, K. B., Owen-Smith, J., & Powell, W. W. (2009). Networks, propinquity and innovation in technological communities. *Administrative Science Quarterly, 54*, 90–122.

Wineman, J. D., Kabo, F. W., & Davis, G. F. (2009). Spatial and social networks in organizational innovation. *Environment and Behavior, 41*, 427–442.

Chapter 3
Extending the Boundaries of Place

Carlos Siordia and Stephen A. Matthews

3.1 Introduction

In *Maphead*, Jennings (2011, p. 41) noted that "the U.S. is the only country in the developed world where a student can go from preschool to graduate school without ever cracking a geography text." We begin with this statement as we believe that demographers, and researchers in other disciplines, interested in spatial analysis can learn a great deal by cracking open a few geography texts and seeking to better understand fundamental spatial concepts such as place, scale, uncertainty, modifiable areal unit problem, spatial autocorrelation, and spatial nonstationarity.[1] Demographers have arrived relatively late to spatial demography, and while it is an exciting field with tremendous research opportunities, it is nevertheless worth reflecting on what we already know about the spatial behavior of the people and the geographical units of analysis (i.e., the spatial contexts) we use in demographic research. Moreover, it is worth reviewing how other disciplines, also late to spatial analysis, define their contextual units of analysis (place); see for example Chaix

[1] The discipline of geography has a long established literature on fundamental spatial concepts, the spatial patterns or spatial dimensions of daily life including constraints on human behavior/activities and in humanistic geography, specifically phenomenological approaches, to the study of meanings associated with, and attachment to, place (for a sample see: Abler et al. 1971; Carlstein et al. 1978; de Smith et al. 2013; Golledge and Stimson 1997; Hägerstrand 1967, 1970; Haggett 1965; Janelle and Goodchild 2011; Tuan 1977). See also Gregory et al. (2009).

C. Siordia
Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: csiordia@gmail.com

S.A. Matthews (✉)
Department of Sociology & Department of Anthropology, State College, PA, USA
e-mail: matthews@psu.edu; sxm27@psu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_3

et al. 2009; Matthews and Yang 2013. The more we examine the wider literature on geospatial data handling and analysis the more evident it should become that how we define place is not a benign choice; the *boundaries of place*, the boundaries of contextual units, matter.[2]

In this chapter we argue that dismissing concerns regarding the definition and measurement of place is analogous to using blinkers (in horse-racing); we might be more focused but we simultaneously introduce blind spots in our research practice. Using geographically defined contexts also requires more spatial thinking. The lack of knowledge of spatial concepts due to limited disciplinary-specific training in this area is one thing[3] but academic "amnesia" more generally is problematic (Gans 1992).[4]

Briefly, in much empirical research that explicitly analyzes individual outcomes (e.g., demographic and health outcomes related to fertility, risky behaviors, morbidity, mortality) within a geographical context tend to adopt naïve views of place. The adoption of naïve views is in part because these views fit our data and analytical models but unfortunately not many of our theories. The geographic boundaries we use to partition and view the world are often problematic, but instead of being viewed as problematic we typically think of them as convenient, and often nothing more. But the convenient, discrete world view confines human action and *exposure to places* by imposing geographically bounded units or levels, and when we adopt this view, we reduce people to actors who cannot escape the "local trap" (Cummins 2007). The local trap is reinforced by our geographic information system (GIS) and multilevel model database structures. We want to be clear that while we criticize GIS and multilevel models this is not a criticism of the database and statistical tools in and of themselves but rather the use of them and the naïve conceptualization of place that is adopted in some studies. We can say this because the people we study have complex lives[5] spanning multiple places (Matthews 2011; Matthews and Yang 2013), and we should note multiple time periods (Kemp 2011). The necessities of life are not confined to a small area fixed in time at one data collection point, but

[2] It is noteworthy that the use of different contextual units can produce different results in analyses (Flowerdew et al. 2008; Mobley et al. 2008; Riva et al. 2008; Roux et al. 2001; Spielman and Yoo 2009; Speilman et al. 2013). Detail discussions on the modifiable areal unit problem (Openshaw 1984) have been given elsewhere (e.g., Fotheringham and Wong 1991; Moon et al. 2005; Root 2012). Readers should be aware that using smaller geographic units may provide better context-measures (Clapp and Wang 2006; Coulton et al. 2011; Hipp 2007), but the call for smaller and smaller nesting units can lead to a non-beneficial degree of reduction (see later example based on Duncan et al. 1961).

[3] As noted in Matthews (Chap. 17, this volume), the instructional opportunities in the area of spatial demography do not necessarily facilitate or promote exposure to the geographical literature.

[4] Academic amnesia refers to the lack of acknowledgment and/or awareness of the work of previous generations of scholars, see Gans 1992.

[5] Individual lives include long commutes, multiple jobs, night shifts, juggling family, daycare, education, social networks and work, finding, preparing, and eating food, practicing one's faith, engaging in leisure activities and exercise, and coping with illness and disease.

exist in a dynamic world that has a history, a present, and a future. The *discrete* view of the world is reinforced by geographically-based contexts and data structures; these ignore the normal day-to-day activity spaces traced out by people as they navigate their complex lives in *continuous* space.[6]

We expand on this issue in the remainder of the chapter. In Sect. 3.2 we will briefly discuss the origins and interest in contextual analysis in the social sciences and describe applications that closely link the scale of the process being studied to the scale of the contextual effect; these tend to be multilevel data structures in society that are based around institutional, organizational, and clear policy levels. Section 3.3 turns to the utilization of geographical contexts by demographers and health researchers. The use of multilevel models in addressing demographic issues has grown rapidly and is now well-established within the field. The question that arises though is whether all the processes we study are adequately captured by the definition of context—or place—that are typically used in our analytical models. While we are critical of some standard model forms, we are particularly excited by emerging methods that seemingly better capture the complexity of human lives. The chapter ends with a synthesis and discussion of several major themes.

3.2 Context in the Social Sciences

Scholarly discussions on the use, appropriateness, and desirability of statistical models that account for contextual effects have existed for several decades or more and their sophistication expanded (e.g., DiPrete and Forristal 1994; Flaherty and Brown 2010; Goldstein 2010; Skrondal and Rabe-Hesketh 2004). Given the original motivation to account for how social context affects human behavior, the social sciences problematized the assumptions of traditional statistical models in the study of multilevel phenomena and began to explore more adequate statistical techniques (Boyd and Iverson 1979; Cronbach et al. 1976; Lindley and Smith 1972). The argument was that classical ordinary least square regressions assume both micro- and macro-level factors were derived from simple random samples, when in effect, many of them were derived from hierarchically structured data (Arnold 1992; Mass and Hox 2004a, b). For example, understanding how a student's socioeconomic status (SES) affected his/her school achievement required that the investigator account for the type of educational system (e.g., private versus public) in which the student participated. If the link between SES and academic

[6] Similarly, the bounded and discrete view of the world reinforces the focus in our analysis on the residential place and also the use of measures of accessibility (potential accessibility) over utilization (revealed accessibility); see Joseph and Phillips 1984.

achievement was found to significantly vary by school sector, then the *classical* statistical approach would not suffice.[7]

More recently, statistical and computing advances have allowed researchers to model individual-level outcomes while accounting for context-level factors by using complex modeling techniques like structural equation modeling and hierarchical modeling (Gelman and Hill 2007; Hox 1995; Kline 2010; Raudenbush and Bryk 2002; Raudenbush et al. 2004; Snijders and Bosker 1999). Although multilevel modeling—as all statistical techniques—has some limitations (Greenland 2000), the advantage of using a multilevel modeling approach was that it allowed the inclusion of a contextual measure by enabling "specification of appropriate error structures, including random intercepts and random coefficients" (Raudenbush 1988, p. 86).[8] These advantages are especially important in nested hierarchies; hierarchies typically found in educational environments (e.g., pupils, within classrooms, within schools, within districts or within sectors) and among organizational structures (Raudenbush and Bryk 1986).

Organizational structures can include social and work-based organizations. For example, an example of the former emerges from the growing body of research is examining the effects of congregation-level characteristics on congregation members. Theoretically, members of religious congregations are posited as being subject to the monitoring and social sanctioning of fellow members (Scheitle and Finke 2008). For example, congregants may be disinclined to espouse a literal view of scripture when surrounded by college-educated members because the surrounding highly-educated members sanction their fellow members' scriptural views (Stroope 2011). Contextual units in this research indicate social interaction and proximity within time and space around shared activities that are plausibly tied to the dependent variable under consideration. More simply, nesting by congregation has theoretical meaning and sociological significance. Multilevel modeling has been used to analyze individual workers nested by organizations and subunits within organizations. Depending on the complexity of an organization, workers are embedded in a variety of contexts that shape particular individual-level factors. Workers may be nested in particular work groups, which may be nested within departments, which are nested in the organization as a whole. As one example, a worker's mood may be positively related to helping behavior. However, a worker's physical proximity to other workers in his or her work

[7] It is worth noting that several of the leading multilevel or hierarchical statistical software packages have their origins in educational and organizational research where many strict nested hierarchies exist: HLM was developed by Bryk & Raudenbush at the University of Chicago while MLN/MLwiN was developed by Goldstein at the Institute of Education, London.

[8] It is beyond the scope of this chapter to discuss how macro-level measures or statistical analysis affect debates over causal mechanisms in multilevel research. Discussion on the relevance of multilevel modeling for identifying casual context effects is given elsewhere (Subramanian 2004). The logic & philosophy of causal inference from statistical analysis is also available elsewhere (Greenland 2011). More detail discussion on the challenges of inferring causality in micro-outcomes (like demography & health) with hierarchical modeling are available (Oakes 2004; Roux 2004) as are calls for improving measurements of group-level constructs when exploring causal mechanisms (Roux 2008).

group (a work group-level effect) may directly affect helping behavior irrespective of mood. Further, mood may more strongly shape helping behavior when a group is more proximal (Hoffman 1997). The work group is the relevant contextual unit and the most theoretically meaningful.

Nesting individuals within contextual levels derived from educational structures, social groupings, and work-place organizational structures are examples of where the higher contextual level has substantive meaning. As we move to consider hierarchies based on geographical units the substantive meaning is less clear cut. Policy makers and social scientists are often interested in policy environments and at some geographical levels this makes sense. The modern nation state provides one of the clearest examples. The nation state has ‘hard’ legal, and typically visible, boundaries. Within these national boundaries all citizens and visitors are bound to the laws and regulations of that contextual level. Within countries there are often several levels that also retain some element of control over local policies and practices; within the U.S. context each state can act independently on a variety of legal and policy matters. Which state one lives in can matter with respect to access and provisioning of welfare and health care services, the level of tax one pays, whether one is mandated to wear a helmet on a motorcycle, whether same-sex couples can marry, and many more things besides.

As we continue to smaller and smaller geographic area the close tie between policy and units of analysis becomes more tenuous. and it is this main issue to which we turn next. Within demography and related fields much of the application of multilevel modeling approaches has used the smaller geographical units to define context or place.

3.3 Demography and Context

Demographers have long sought to understand how social context shapes individuals’ life experiences.[9] Demographers and others have written of a ‘tidal wave’ or ‘explosion’ in neighborhood effects studies, particularly studies focusing on the structural conditions of neighborhoods such as poverty and race/ethnic composition (Dietz 2002; Entwisle 2007; Pebley and Sastry 2004; Roosa et al. 2003; Roux and Mair 2010). This explosion has been driven by the availability of new analytical methods, software and data that facilitates the analysis and integration of data on people and places.[10] New types of geospatial data and emerging spatial statistical

[9] Over a century ago, Ravenstein (1876) noted the link between birthplace and migratory behavior. In the early part of the twentieth century Park (1926, p. 18) wrote that the use of statistical methodologies in social science were important only “because social relations are so frequently & so inevitably correlated with spatial relations.”

[10] It is beyond the scope of this paper but we note that there is a growing realization that there is need for both the development of new and the validation of existing relevant place-level measures (e.g., built environment measures). Similarly, there is a need for research on identifying sources of spatial uncertainty (i.e., inaccuracy or instability) in both existing & new kinds of data demographers will utilize.

methods are both helping promote spatially informed demographic research with the United States and across the globe.

The re-emergence of spatial demography has been robust (Voss 2007; Matthews and Parker 2013; and noted by several contributors to this edited collection). Today many research and policy questions faced by demographers require the analysis of complex patterns of interrelated social, behavioral, economic, and environmental phenomena and both spatial thinking and spatial analytical perspectives have an important role to play in addressing these types of questions (Castro 2007). Indeed, as Entwisle (2007)) noted in her Presidential address to the Population Association of America (PAA) cutting-edge demographic research will increasingly depend on the collection and analysis, and the integration, of both individual- and contextual-level data across a wide range of spatial and temporal scales.

Theoretically, the potential relevance of contexts or 'higher-levels' to individual and family outcomes is widely recognized. Indeed across many academic fields there has been a long-standing interest in the effects of broader social and geographic contexts on human behavior (Gieryn 2000). Bronfenbrenner's (1979) ecological perspective and typology of multiple, overlapping, individual, and environmental contexts is perhaps among the best-known general contextual frameworks. More specific to geographical areas as context researchers set individuals within hierarchies of urban social space (Chombart de Lauwe 1960), nested hierarchies of place (Suttles 1972), and Jacobs (1961) levels of neighborhood (1961). These are classic studies, all written 30+ years ago. More recently, across the social and health sciences scholars have repeatedly argued that research practice is increasingly interested in the associations between individual-level outcomes and contextual effects. That is, there is now considerable interest in more *distal*, *macro*, and *fundamental causes* and the processes operating at multiple scales (Booth and Crouder 2001; Galea 2007; Kawachi and Berkman 2003; Leventhal and Brooks-Gunn 2000; Link and Phelan 1995; Roosa et al. 2003; Schonkoff and Phillips/NAP 2000; Taylor et al. 1997).

While several conceptual models and frameworks exist, our critical review of these conceptual models and frameworks is that they are vague on what they mean by 'place;' and specifically which geographical units are relevance to the contextual process and the outcome of interest.

Paraphrasing Galster (2001) it seems that in our conceptual models 'everyone seems to know what place is' but if we are honest there is little clarity on what *levels* of analysis can and should be used in our empirical models. We say *levels* as we believe that there is more to places than people think. To borrow an analogy from the movie Shrek, replacing the word ogre with place, then "Places are like an onion. . . . Onions have layers. Places have layers" (Shrek 2001). This multi-layered perspective on place is reflected in several conceptual models (National Academy of Sciences 2006). While many of these frameworks make frequent reference to distal factors such as social structure, social environments, social/political/economic conditions, policy, and environmental resources and constraints, if reference

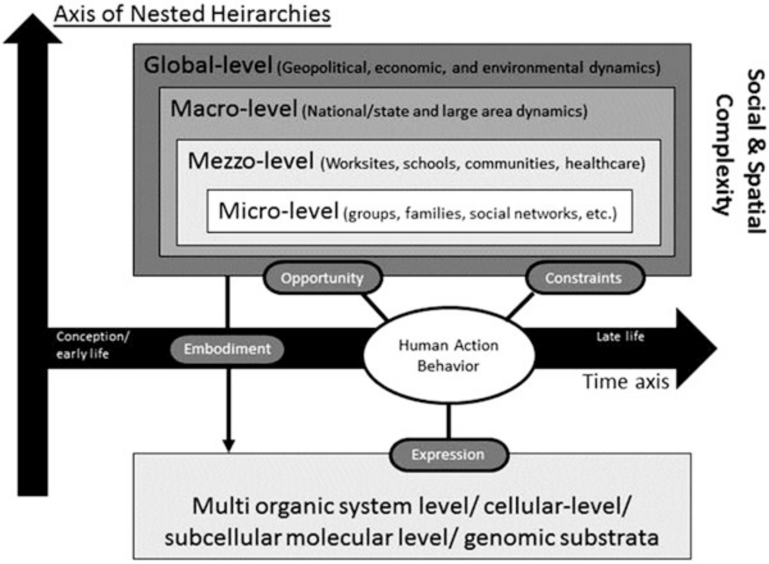

Fig. 3.1 The society-behavior-biology nexus (Adapted from Figure 1. Glass and McAtee 2006)

is made to a level or a context it is often neighborhoods or communities, where the actual definition of these is unspecified.[11]

The framework proposed by Glass and McAtee (2006) is among the more specific about nested levels (see Fig. 3.1). Glass and McAtee focus on the nesting of supra-contextual levels, labeled as *micro*, *mezzo*, *macro*, and *global*; and within each of these they identify specific examples of analytical levels of interest. Interestingly, these analytical levels are a mix of social, organizational, and geographical units or contexts. In the authors' defense, and as we have seen in Sect. 3.2, these reflect the main types of analytical units most frequently used in multilevel research. However, the overall framing of levels is one of a nested structure. Note that some levels are more clearly geographically bounded places or as explicit types of places (worksite, schools). Thus, for the most part, these places are often non-nested organizationally and may even be non-overlapping geographically (e.g., residential tract, workplace, school district).

We posit that nesting by school is *not* similar to nesting by geographical polygon based on place of residence.

Demography is defined as "the statistical study of human populations" and has long been associated with careful measurement and analysis. In a variety of

[11] Clapp and Wang (2006, p. 260) suggest that a lack in definitional precision but wide use of the term "neighborhood" may occur because "the ordinary language definition is considered so compelling as to require little elaboration." While using the label neighborhood may provide a perception of conceptual power, arguably modest scientific insight is afforded by the ambiguous use of the term.

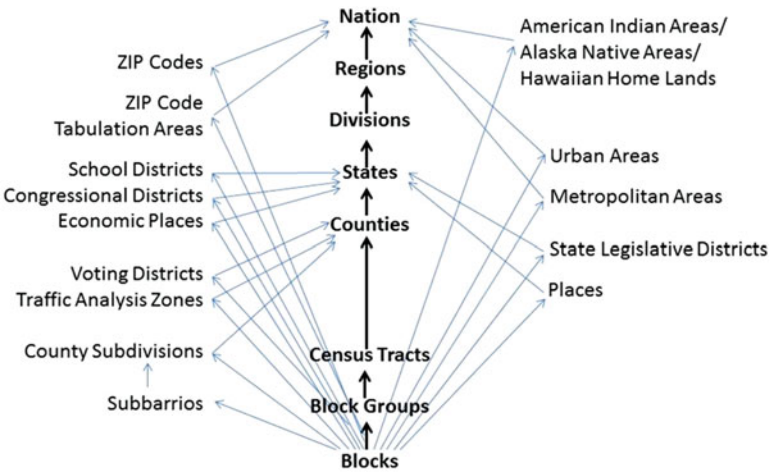

Fig. 3.2 Some of the nested and non-nested geographical hierarchies of the U.S. Census

sub-fields demographers focus on collecting data and studying transitions, sequences of transitions, and changes in behavior from one state to another (e.g., school to work, unemployed to employed, work to retirement, single to married, a non-smoker to smoker) and on the timing, duration, and sequencing of events (e.g., first intercourse, first birth, first drug use, disease diagnosis, and survival after diagnosis). Demographic research on transitions and change requires a careful attention to an understanding of timing issues and the relation between time intervals and the processes of change being studied. Thus demographers interested in studying processes around marriage and fertility are reasonably happy with annual data sequences, their colleagues studying labor markets would prefer monthly and seasonal data, and yet others looking at household divisions of labor can draw on data sequences collected by the minute.[12] The main point here is that as a field, demography has been acutely sensitive to issue surrounding the measurement of time; in contrast the measurement of place has received scant attention.

The demographer frequently falls back on census geographies as the favorite units of analysis, but we maintain that they rarely ask questions regarding the assumptions being made due to the selection a specific unit or census geography. Figure 3.2 is a representation of the geographical hierarchies used by the U.S. Census Bureau and agencies across the federal government. This diagram highlights the well-known levels and hierarchies and also, the non-nested hierarchies that exist. What is also visible, but less well-known, is that none of the administrative levels above the census block (the levels within the strict eight-level hierarchy in the central column) can be combined to form aggregate areas necessarily matching those census geographies shown on the left and right of the

[12] Our ability today to collect data at increasingly finer temporal scales, and aggregate up as needed, provides analytical flexibility never before seen.

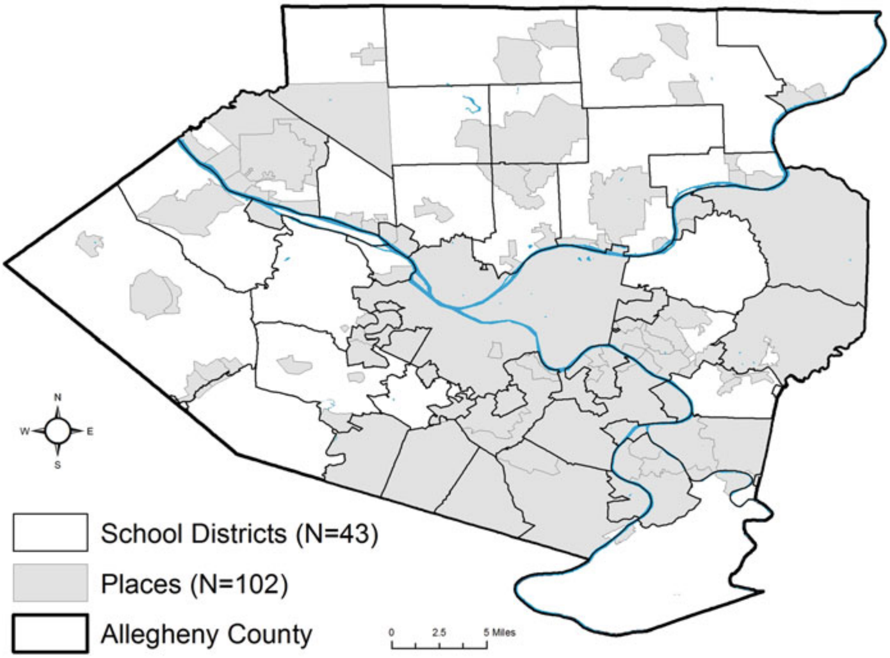

Fig. 3.3 Example of selected geographies in Allegheny County, Pennsylvania, United States

diagram (i.e., places are not necessarily embedded within a single county, a tract is not necessarily embedded within a ZIP code or a school district, and school districts don't have to align with counties).[13] Figure 3.3 shows places and school district boundaries within Allegheny County (Pittsburgh), Pennsylvania. Places do not provide complete spatial coverage. Both places and school district boundaries do not always align, vary widely in size and shape, and follow complex patterns that make the link between theory and measurement challenging. Not visible in Fig. 3.2 is the fact that some census geographies do not provide complete spatial coverage (e.g., places and urban areas). The lack of spatial correspondence and the lack of complete spatial coverage are both important issues for any researcher examining contextual effects.[14]

Returning to Glass and McAtee we can focus on key variables—dimensions of interest—that researchers use as attributes of place (Fig. 3.4).[15] All of these

[13] In some states school districts may align with counties but in others they do not. That is, some pairs of levels within the census hierarchy do not have a consistent relationship across the whole country.

[14] There is not space in this chapter to discuss boundary changes over time or that in some states but this is another important wrinkle.

[15] We are using this diagram from Glass and McAtee out of context. Our interest is only on discussing the main dimensions of interest.

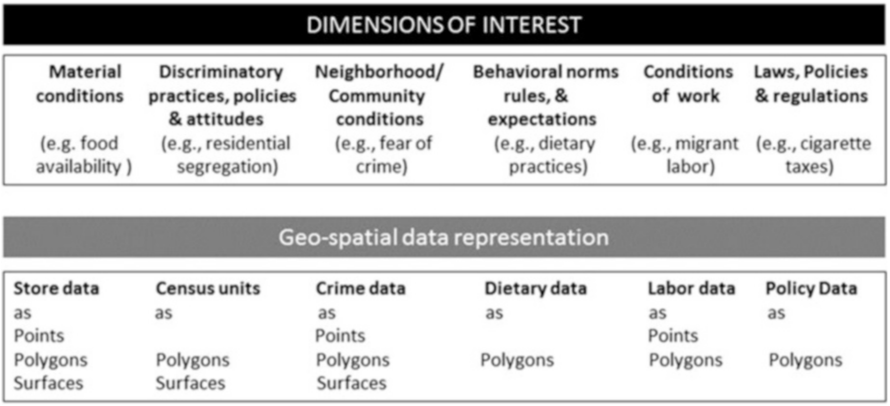

Fig. 3.4 Dimensions of interest about place and their representation in GIS databases (Adapted from Figure 2. Glass and McAtee 2006)

examples are substantively relevant variables but several issues are worth closer scrutiny. At the most fundamental the question we ought to ask is: *At what contextual level is X measured?* (where, X equals food availability, residential segregation, crime, dietary practices, migrant labor, or cigarette tax).

If we choose a single definition of place another important question related to how place "gets under the skin" is *Are all dimensions of interest and the processes and mechanisms they represent operating at this level?* While some of the examples of Fig. 3.4 do have connections to policy relevant levels (e.g., cigarette taxes) all other dimensions of interest have no obvious tie to a specific level. Indeed the lower half of Fig. 3.4 indicates how these dimensions/indicator variables can be measured and represented in a GIS. Moreover, for some of these variables there are multiple measures that could be used; for example, Massey and Denton (1988) describe 20 different indexes—grouped into five key dimensions of segregation: evenness, exposure, concentration, centralization, and clustering.

Race/ethnic segregation is based on a single common source—the U.S. Census Bureau—but for some variables there are a wide variety of sources. For example, consider the different ways in which the food environments or food availability can be measured (see the National Collaborative on Childhood Obesity Research (NCCOR) Measures Registry at URL: http://nccor.org/projects/measures/index.php). If the data on food stores comes as geocoded points then the analyst can aggregate to generate density scores or measures of the diversity of different store types for geographic areas (polygons) or generate density and/or accessibility surfaces. In theory point data provides flexibility but also the aggregation of such data and manipulation to match to areal units also can lead to different 'scores' for input into a model. Fifty years ago, Duncan et al. (1961, page 35, Table 3.1) provide a simple but effective illustration of the problem vis-à-vis calculating a simple variable such as population density around a specific known location or address (within the tract #550).

Table 3.1 Illustrative population densities of various areal units, 1950

Geographic unit	Land area (sq. miles)	Population density
Chicago, census tract (#550)	0.02	91,300
Chicago, community area (#35)	1.62	48,500
Chicago, city	207.50	17,450
Chicago, urbanized area	638.00	7,713
Chicago, metropolitan area	1,184.20	4,283
Chicago, standard metropolitan area	3,617.00	1,519
East North Central Division	244,867.00	124
Continental United States	2,974,726.00	51

Source: Duncan et al. (1961)

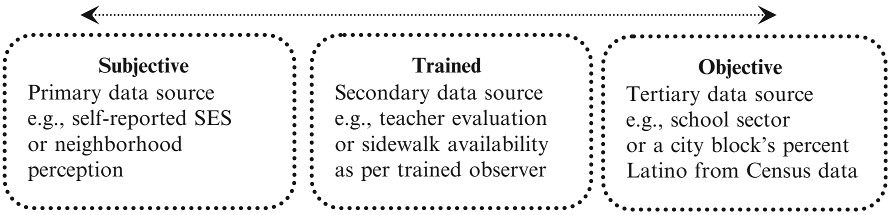

Fig. 3.5 Continuum of contextual information (Booth et al. 2005)

As these examples indicate there are a wide range of information sources and measures available to a researcher interested in the attributes of place. Figure 3.5, adapted from Booth et al. (2005), displays the contextual information resource continuum. On the left-end of the continuum, we have "subjective" information sources—which include items like a person's neighborhood perception (e.g., fear of crime). Neighborhood perception can be an outcome measure at the individual level but survey variables can also be used to derive aggregate response variables (e.g., the average fear of crime derived from all those individuals within a shared group or geographical contexts. While "trained" information resources include measures collected by trained observers (e.g., neighborhood audits) or from independent secondary sources (e.g., teacher evaluations). On the right, "objective" information sources are those most commonly used by quantitative demographers (e.g., data from the U.S. Census Bureau).

Figure 3.6 summarizes the basic problem of mismatch between the on the one hand the processes and mechanisms we are interested in and the unit of analysis used to define context. If the process we are studying is small but our unit of analysis is a much larger place, then the aggregated data we have misses capturing the local nuances (the case of the small child in the big chair). Conversely, if the process we are studying is large but our unit of analysis is small, then we end up with information at too small a unit. Not all context units of analysis are equal in their level of meaning and usefulness (Lupton and Kneale 2012).

As we stated earlier, it is important to note that in many multilevel models within demography and social science there is a tendency to use relatively small geographic units to define context (e.g., census blockgroups and census tracts) and in

Fig. 3.6 Measuring processes at an appropriate scale (Photo: Courtesy of Stephen Matthews (May 5, 2012) San Francisco Exploratorium)

such models to utilize variables such as the unemployment rate or a measure of race/ethnic segregation are used. However, local unemployment rates and race/ethnic segregation do not occur in a vacuum. Census tracts are embedded in nested and non-nested larger geographical regions, and the drivers of unemployment and race/ethnic segregation can emerge from macro-level factors—housing and labor markets—that may operating across county or multi-county areas, which are in turn embedded within regional and national trends. That is, contextual levels in multilevel models are treated as discrete and the processes we seek to understand are modeled as discontinuous processes. As Fotheringham et al. (2003, p. 19) conclude "it is assumed that the process is modified in exactly the same way throughout a particular spatial unit but that the process is modified in a different way as soon as the boundary of that spatial unit is reached (2003, p. 19)." In this, the standard multilevel conceptualization of the real world, human behavior matters most within a specific contextual unit. This can be true of organizational and group contexts but is less likely the case in geographical contexts. For example, if one were to generate a tract-level map of the race/ethnic structure of Atlanta, focusing on the percent African American, one would immediately recognize that the 'scale' of race/ethnic segregation is a function of processes beyond the scale of the census tract (Lee et al. 2008; Reardon et al. 2008). We note that more complex multilevel models can handle higher-level embeddedness and both nested (e.g., 3-level models

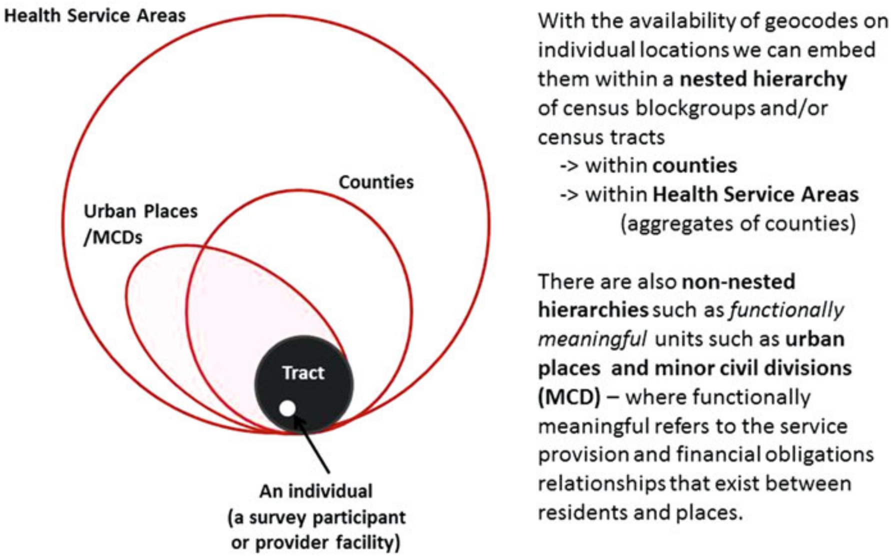

Fig. 3.7 The complexity of nested and non-nested contextual levels

such as individuals within tracts within health service areas, see Fig. 3.7) and non-nested multiple membership models (e.g., individuals within tracts, within urban places within counties); see Browne et al. (2001). The examples used in Fig. 3.7 are also meant to make the claim that certain contextual levels (places) are functionally meaningful (Flowerdew et al. 2008).

Figure 3.8 displays two- and three-dimension images of how an individual can be nested within multiple spatial units or places (Matthews 2011).[16] An individual during the course of routine activities may follow many pathways and these can differ in purpose, scale, duration, and frequency. Frequency over pathways is represented by the width of the arrows. For example, the widest arrow signals the pathway most frequency used by an individual, the journey to work. Other trips and activities take the individual to other places, each potentially defined by different units of geography. As such, the use of any one representation of place, and lack of attention to the temporal duration within places, may mean that researchers over- or under-represent actual exposure to the right place (Chaix et al. 2013; Rainham et al. 2010; Zenk et al. 2011). Researchers in physical activity research, more so than in demography, are beginning to examine the feasibility of utilizing GPS data to create person-specific activity space measures and to use summary measures (convex hull, standard deviational ellipses, buffers, and kernel density estimation) to define environmental or place-specific exposure (see Matthews and Yang 2013).

[16] Note that Fig. 3.8 uses the same geographical levels as Fig. 3.7.

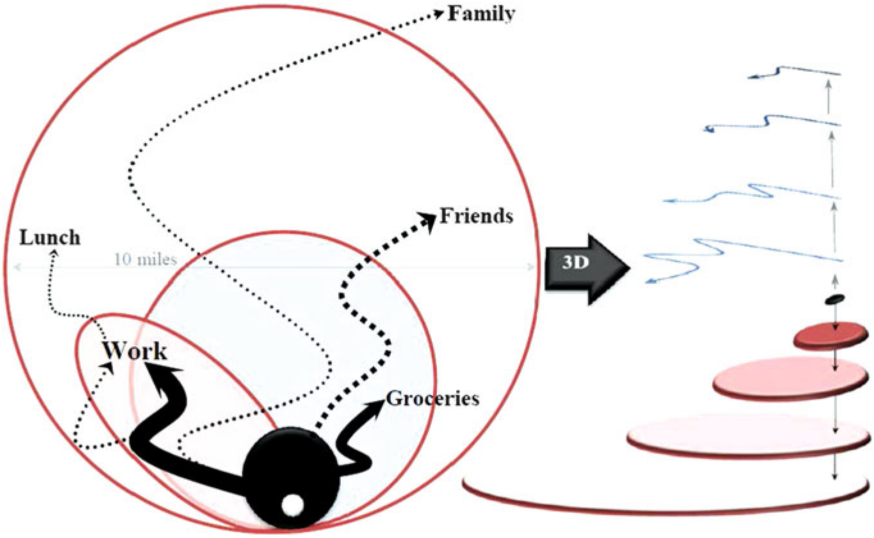

Fig. 3.8 Person-based pathways through nested and non-nested contextual levels

3.4 Conclusions

In this chapter we have argued that the definition of place is not benign in our analytical models. We focused on the use of multilevel models in demographic and health scholarship as this is now a well-established analytical tool within the field. The popularity of multilevel models is well deserved as they offer statistical advantages over conventional single-level approaches, and are especially relevant in detecting contextual effects. The conventional two-level multilevel model—situating an individual within a single context—is appropriate when the context is defined within rigid hierarchical structures (e.g., pupils in classrooms, individuals within a workplace, an organization or a group; individuals within families) but much less so when the higher level context is a geographical unit of analysis. Of course there are exceptions such as the nation state and other sub-national places for which public policy decisions apply uniformly. These sub-national places would include in the U.S. context states as policies can vary across states but are uniformly applied within them.

If an analytical model is seeking to examine pathways and exposures linking people to place then it would seem that several functionally meaningful units reinforce this relationship. Counties, minor civil divisions, local governments (census places and urban places), and school districts are in many instances highly salient—functionally relevant—contexts. It is within these contexts that both service provision (revenue spending) and financial obligations (taxation) relationships exist between individual residents and the place of residence. We have argued that other, smaller, contextual units of analysis are more problematic and in this set we include census blockgroups, tracts, and ZIP codes. To quote form one of the defining text in geography by Abler et al. (1971, p. 566) “Areal units are particularly sacred once they have been established even though they later may become serious obstacles to the solutions of contemporary problems.” The obstacle that small areas provide, specifically census blockgroups and census tracts, is that they are inadequate contextual units for examining *how place gets in to people*. In general, these units are too small and place unrealistic boundaries around the lives of people in contemporary society. Small areas may be appropriate for the study of certain populations such as young children and relatively immobile elderly based on the assumption that immobility glues them to these confined places. However, even the lives of the less mobile are linked to multiple places beyond the local (e.g., not every census tract has a childcare center, an elementary school, a doctor, a park).[17]

Our intent in this chapter has been to promote more clear thinking on the processes and mechanisms linking people to place. Current practice and the emphasis on small areas (tracts or neighborhoods/communities defined as tracts) illustrates the mismatch between our understanding of the spatial and temporal scales of human behavior and scales of analysis (see Matthews and Yang 2013). Interestingly in demography-related disciplines the focus of exposure-type studies has been on census derived neighborhoods coupled with a heavy focus on census derived variables. It is worth noting that this census area and census variable approach is in sharp contrast to some of the research in other fields (e.g., public health, and especially physical activity research), who are less dependent on ‘social’ data and are increasingly using point-based data on built environment features and/or raster data on physical environment attributes to generate measures for egocentric buffers, sometimes at multiple scales, and for activity spaces (Chaix et al. 2009). That is, researchers in other fields, also examining the relations between people and place, have not always been tied to administrative units and in general been more creative in both their definition of place and in measuring the attributes of place.[18]

We join others (including Chaix et al. 2009; Cummins et al. 2007; Kwan 2012; Matthews et al. 2009; Roux 2004, 2007, 2008) and call for greater discussion of the meaning and/or relevance of context measures. Place effect investigations weakly operationalize the meaning of contextual measures and consequently provide

[17] Parenthetically the absence of these local resources in the residential tract or context does not necessarily make these areas a childcare desert, a school desert, a medical desert, or a park desert.

[18] Demographers who have been used to relying on the Summary Tape Files or Summary Files (the census long-form) data to define neighborhood attributes now have to use the American Community Survey 5-year estimates. As the margin of error can be high for both the population & the housing variables from the ACS it might be prudent to start thinking about using larger contextual units and tapping in to alternative sources of data on the social, built, and physical environment.

The American Community Survey (ACS) implies a focus on “community;” a phrase that along with “neighborhood” has been employed loosely across the social & health sciences.

modest or no discussion of the mechanisms linking the particular contextual unit to the individual outcome of interest. Place effects research must be based on more solid theoretical foundations.

We do not believe there is one appropriate definition of context that is applicable to all outcomes. There is no Holy Grail in contextual effects research based on geographical areas. There are, however, several geographical contexts that would appear to be more relevant than others in shaping micro-level behaviors. We have described them as the functionally relevant contexts and we believe there is potentially a rich seam to mine at these levels.

References

Abler, R., Adams, J. S., & Gould, P. (1971). *Spatial organization: The geographer's view of the world*. Englewood Cliffs: Prentice Hall International.

Arnold, C. L. (1992). An introduction to hierarchical linear models. *Measurement and evaluation in counseling development, 25*(2), 58–90.

Booth, A., & Crouter, A. C. (Eds.). (2001). *Does it take a village? Community effects on children adolescents, and families*. Mahwah: Lawrence Erlbaum.

Booth, K. M., Pinkston, M. M., & Poston, W. S. C. (2005). Obesity and the built environment. *Journal of the American Dietetic Association, 105*(5), S110–S1116.

Boyd, L. H., & Iverson, G. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont: Wadsworth Publishing.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling, 1*(2), 102–124.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Carlstein, T., Parks, D., & Thrift, N. (Eds.). (1978). *Timing space and spacing time: Human activity and time geography* (Vol. 2). New York: Wiley.

Castro, M. (2007). Spatial demography: An opportunity to improve policy making at diverse decision levels. *Population Research and Policy Review, 265*(5–6), 477–509.

Chaix, B., Merlo, J., Evans, D., Leal, C., & Havard, S. (2009). Neighborhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio & Brodeur. *Social Science and Medicine, 69*(9), 1306–1310.

Chaix, B., Méline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C., et al. (2013). GPS tracking in neighborhood and health studies: A step forward for environmental exposure assessment, a step backward for causal inference? *Health and Place, 21*, 46–51.

Chombart de Lauwe, P. H. (1960). L'evolution des besoins et la conception dynamiques de la famille. *Revue Française de Sociologie,* 403–425.

Clapp, J. M., & Wang, Y. (2006). Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban Economics, 59*(2), 259–284.

Coulton, C. J., Chan, T., & Mikelbank, K. (2011). Finding place in community change initiatives: Using GIS to uncover residential perceptions of their neighborhoods. *Journal of Community Practice, 19*(1), 10–28.

Cronbach, L. J., Deken, J. E., & Webb, N. (1976). *Research on classrooms and schools: Formulation of questions, design and analysis* (Occasional Paper, Stanford Evaluation Consortium, Stanford University). Stanford: Stanford Evaluation Consortium.

Cummins, S. (2007). Commentary: Investigating neighbourhood effects on health—Avoiding the 'Local Trap'. *International Journal of Epidemiology, 36*(2), 355–357.

Cummins, S., Curtis, S., Diez-Roux, A. V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Social Science and Medicine, 65*(9), 1825–1838.

de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2013). *Geospatial analysis* (4th ed.). Leicester: The Winchelsea Press, Troubador Publishing Limited. Available at http://spatialanalysisonline.com/

Dietz, R. D. (2002). The estimation of neighborhood effects in the social sciences: And interdisciplinary approach. *Social Science Research, 31*(4), 539–575.

DiPrete, T. A., & Forristal, J. D. (1994). Multilevel models: Methods and substance. *Annual Review of Sociology, 20*, 331–357.

Duncan, O. D., Cuzzort, R. P., & Duncan, B. (1961). *Statistical geography: Problems in analyzing areal data*. Glencoe, IL : The Free Press.

Entwisle, B. (2007). Putting people into place. *Demography, 444*, 687–703.

Flaherty, J., & Brown, R. B. (2010). A multilevel systemic model of community attachment: Assessing the relative importance of the community and individual levels. *The American Journal of Sociology, 116*(2), 503–542.

Flowerdew, R., Manley, D. J., & Sabel, C. E. (2008). Neighbourhood effects on health: Does it matter where you draw the boundaries? *Social Science and Medicine, 66*(6), 1241–1255.

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A, 23*(7), 1025–1044.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2003). *Geographically weighted regression: The analysis of spatially varying relationship*. New York: Wiley.

Galea, S. (Ed.). (2007). *Macrosocial determinants of population health*. New York: Springer.

Galster, G. C. (2001). On the nature of neighborhoods. *Urban Studies, 38*(12), 2111–2124.

Gans, H. (1992). Sociological amnesia: The noncumulation of normal social science. *Sociological Forum, 7*(4), 701–710.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gieryn, T. (2000). A space for place in sociology. *Annual Review of Sociology, 26*, 463–496.

Glass, T. A., & McAtee, M. J. (2006). Behavioral science at the crossroads in public health: Extending horizons, envisioning the future. *Social Science and Medicine, 62*(7), 1650–1671.

Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). New York: Wiley.

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: A geographical perspective*. New York: Guilford Press.

Greenland, S. (2000). Principles of multilevel modeling. *International Journal of Epidemiology, 29*, 158–167.

Greenland, S. (2011). The logic and philosophy of causal inference: A statistical perspective. In D. M. Gabbay, P. Thagard, & J. Woods (Eds.), *The handbook of the philosophy of science* (Philosophy of statistics, Vol. 7). New York: Elsevier.

Gregory, D., Johnston, R., Pratt, G., Watts, M., & Whatmore, S. (Eds.). (2009). *The dictionary of human geography*. New York: Wiley.

Hägerstrand, T. (1967). *Innovation diffusion as a spatial process* (A. Pred, Trans.). Chicago: University of Chicago Press.

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science, 24*(1), 7–21.

Haggett, P. (1965). *Locational analysis in human geography*. London: Edward Arnold.

Hipp, J. R. (2007). Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review, 72*(5), 659–680.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*(6), 723–744.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT- Publikaties.

Jacobs, J. M. (1961). *The death and life of great American cities*. New York: Random House Inc.

Janelle, D. G., & Goodchild, M. F. (2011). Concepts, principles, tools and challenges in spatially integrated social science. In T. L. Nyerges, H. Couclelis, & R. McMaster (Eds.), *The SAGE handbook of GIS and society*. Thousand Oaks: Sage.

Jennings, K. (2011). *Maphead: Charting the wide, weird world of geography wonks*. New York: Scribner.

Joseph, A. E., & Phillips, D. R. (1984). *Accessibility and utilization: Geographical perspectives on health care delivery*. Thousand Oaks: Sage.

Kawachi, I., & Berkman, L. F. (2003). *Neighborhoods and health*. New York: Oxford University Press.

Kline, R. B. (2010). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kemp, S. P. (2011). Place, history, memory: Thinking time within place. In L. M. Burton et al. (Eds.), *Communities, neighborhoods, and health: Expanding the boundaries of place* (pp. 3–19). New York: Springer.

Kwan, M.-P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers, 102*(5), 958–968.

Lee, B. A., Reardon, S. F., Firebauch, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the census tract: patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review, 73*, 766–791.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309–337.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B Methodological, 34*(1), 1–41.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior, 35*(Extra issue), 80–94.

Lupton, R., & Kneale, D. (2012). Theorising and measuring place in neighborhood effects research: The example of teenage parenthood in England. In M. van Ham, D. Manley, N. Vailey, et al. (Eds.), *Neighborhood effects research* (pp. 121–145). New York: Springer.

Mass, C. J. M., & Hox, J. J. (2004a). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127–137.

Mass, C. J. M., & Hox, J. J. (2004b). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis, 46* (3), 427–440.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*, 281–315.

Matthews, S. A. (2011). Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods. In L. M. Burton, S. Kemp, M. Leung, S. A. Matthews, & D. Takeuchi (Eds.), *Communities, neighborhoods, and health: Expanding the boundaries of place*. New York: Springer.

Matthews, S. A. (2012a). Thinking about place, spatial behavior, and spatial processes in childhood obesity. *American Journal of Preventive Medicine, 42*(5), 516–520.

Matthews, S. A. (2012b). Frontiers in spatial demography and population geography. In Presentation given at the annual meetings of the *Association of American Geographers*. New York.

Matthews, S. A., & Parker, D. M. (2013). Progress in spatial demography. *Demographic Research, 28*(10), 271–312.

Matthews, S. A., & Yang, T.-C. (2013). Spatial Polygamy and Contextual Exposures (SPACEs): Promoting activity space approaches in research on place and health. *American Behavioral Scientist, 57*(8), 1057–1081.

Matthews, S. A., Moudon, A. V., & Daniel, M. (2009). Using Geographic Information Systems (GIS) for enhancing research relevant to policy on diet, physical activity, and weight. *American Journal of Preventive Medicine, 36*(4S), 171–176.

Mobley, L. R., Kuo, T., & Andrews, L. S. (2008). How sensitive are multilevel regression findings to defined area of context? A case study of mammography use in California. *Medical Care Research and Review, 65*(3), 315–337.

Moon, G., Subramanian, S. V., Jones, K., Duncan, C., & Twigg, L. (2005). Area-based studies and the evaluation of multilevel influences on health outcomes. In A. Bowling & S. Ebrahim (Eds.), *Handbook of health research methods*. New York: Open University Press.

National Academy of Sciences. (2006). *Examining the health disparities research plan of the National Institutes of Health: Unfinished business*. Institutes of Medicine, Washington, DC: National Academies Press.

Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science and Medicine, 58*(10), 1929–1952.

Openshaw, S. (1984). *The modifiable areal unit problem*. Norwich: Geo Books.

Park, R. E. (1926). The urban community as a spatial pattern and a moral order. In E. W. Burgess (Ed.), *The urban community*. Chicago: University of Chicago Press.

Pebley, A. R., & Sastry, N. (2004). Neighborhoods, poverty, and children's well-being. In K. M. Neckerman (Ed.), *Social inequality* (pp. 119–145). New York: Russell Sage Foundation.

Rainham, D., McDowell, I., Krewski, D., & Sawada, M. (2010). Conceptualizing the healthscape: Contributions from time geography, location technologies and spatial ecology to place and health research. *Social Science and Medicine, 70*(5), 668–676.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics, 13*(2), 85–116.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1–17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. Jr. (2004). *HLM 6 for Windows* [Computer software]. Lincolnwood: Scientific Software International, Inc.

Ravenstein, E. G. (1876). *The birthplaces of the people and the laws of migration*. London, UK: Trübner & Company.

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography, 45* (3), 489–514.

Riva, M., Apparicio, P., Gauvin, L., & Brodeur, J.-M. (2008). Establishing the soundness of administrative spatial units for operationalizing the active living potential of residential environments; an exemplar for designing optimal zones. *International Journal of Health Geographics, 7*(1), 43.

Roosa, M. W., Jones, S., Tein, J.-Y., & Cree, W. (2003). Prevention science and neighborhood influences on low-income children's development: Theoretical and methodological issues. *American Journal of Community Psychology, 31*(1/2), 55–72.

Root, E. D. (2012). Moving neighborhood and health research forward: Using geographic methods to examine the role of spatial scale in neighborhood effects on health. *Annals of the Association of American Geographers*, available on-line on April 3, 2012. doi:10.1080/00045608.2012.659621

Roux, A. V. D. (2004). Estimating neighborhood health effects: The challenges of causal inference in a complex world. *Social Science and Medicine, 58*(10), 1953–1960.

Roux, A. V. D. (2007). Neighborhoods and health: Where are we and where do we go from here? *Revue d'Épidémiologie et de Santé Publique, 55*(1), 13–21.

Roux, A. V. D. (2008). Next steps in understanding the multilevel determinants of health. *Journal of Epidemiology and Community Health, 62*(11), 957–959.

Roux, A. V. D., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences, 1186*(1), 125–145.

Roux, A. V. D., Kiefe, C. I., Jacobs, D. R., Hann, M., Jackson, S. A., et al. (2001). Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology, 11*(6), 395–405.
Scheitle, C. P., & Finke, R. (2008). Maximizing congregational resources: Selection versus production. *Social Science Research, 37*(3), 815–827.
Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Committee on Integrating the Science of Early Childhood Development, Board on Children, Youth, and Families. Washington, DC: The National Academies Press.
Shrek. (2001). *Dreamworks pictures*. Universal City, CA.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton: Chapman and Hall/CRC.
Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modelling*. London: Sage.
Spielman, S. E., & Yoo, E. H. (2009). The spatial dimensions of neighborhood effects. *Social Science and Medicine, 68*(6), 1098–1105.
Spielman, S. E., Yoo, E.-H., & Linkletter, C. (2013). Neighborhood contexts, health, and behavior: Understanding the role of scale and residential sorting. *Environment and Planning B: Planning and Design, 40*(3), 489–506.
Stroope, S. (2011). Education and religion: Individual, congregational, and cross-level interaction effects on biblical literalism. *Social Science Research, 40*(6), 1478–1493.
Subramanian, S. V. (2004). The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Social Science and Medicine, 58*(10), 1961–1967.
Suttles, G. D. (1972). *The social construction of communities* (Studies of urban society). Chicago: University of Chicago Press.
Taylor, S. E., Repetti, R. L., & Seeman, T. (1997). Health psychology: What is an unhealthy environment and how does it get under the skin? *Annual Review of Psychology, 48*(1), 411–447.
Tuan, Y.-F. (1977). *Space and place: The perspective of experience*. Minneapolis: University of Minnesota Press.
Voss, P. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*(5), 457–476.
Zenk, S. N., Schulz, A. J., Matthews, S. A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., et al. (2011). Activity space environment and eating and physical activity behaviors: A pilot study. *Health and Place, 17*(5), 1150–1161.

Chapter 4
Using Place-and Person-Based Interventions to Measure Neighborhood Effects

Noli Brazil

4.1 Introduction

Although the study of spatial variation in individual outcomes has a long and rich tradition, only recently have researchers focused on investigating how local and spatial contexts, or more specifically neighborhoods, affect the well-being of residents. Attempts at understanding the influence of a neighborhood on individual outcomes have spawned a vast literature on neighborhood effects in the United States and other developed countries. Earlier research in the field established broad social theories of neighborhood influence and empirically tested crude measures of community-level factors that demonstrated correlated links between place and individual well-being. Both modes of research were vital in advancing the field from its infancy, but as the literature matured, several challenges that require advancements in the conceptualization and methodological testing of neighborhood effects emerged, such as defining the appropriate scale of neighborhood, quantifying exposure, measuring appropriate neighborhood characteristics, and modeling heterogeneity (Raudenbush and Sampson 1999; Galster 2008; Wodtke et al. 2011; Harding et al. 2011).

Researchers have expressed significant concern with two specific challenges: identifying mechanisms of neighborhood effects and developing methods that minimize endogeneity. Although significant progress has been made to overcome these challenges, work in these two streams of research have largely been done independently of one another. One set of research establishes broad theories of neighborhood effects that have been difficult to empirically test while the other set develops methods that deemphasize space and trade testable hypotheses for internal

N. Brazil (✉)
Center for Research on Inequalities and the Life Course, Yale University,
New Haven, CT, USA
e-mail: noli.brazil@yale.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_4

validity. A specific trend in the field that underscores the disconnect is the use of randomized housing mobility programs to measure neighborhood effects. While these interventions minimize the selection problems plaguing earlier studies by relying on randomization to place individuals into neighborhoods of varying quality, they are devoid of any theoretical grounding.

Researchers will continue to support experimental approaches to measuring neighborhood effects given their methodological advantages over observation-based methods. The challenge is to mesh results from these experiment-based strategies with the broad theories introduced by earlier neighborhood scholars. The purpose of this chapter is to develop a conceptual framework that links experiment-based methods with testable middle-range theories of neighborhood effects. Specifically, the framework organizes experiment-based neighborhood effects into a typology, person- and place-based, and further decomposes the effects into their natural direct and indirect components. The typology sheds some light on how researchers might think about both people and places by developing the theoretical foundations and assumptions that underlie these alternative approaches. The decomposition unbundles the average neighborhood effect into components that capture the various mechanisms an experimental intervention works through to influence individual outcomes, making the comparison of person- and place-based effects more tractable and allowing the researcher to parse out the true neighborhood effect from contaminated parts.

4.2 Middle-Range Theories and Selection Bias

Although the philosophical inquiry into the role of neighborhoods in shaping individual outcomes dates back to the early twentieth century (Galpin 1915; Park and Burgess 1925), it wasn't until much later, through the pioneering work of scholars such as Wilson (1987), Jencks and Mayer (1990) and urban sociologists of the classical Chicago school (Sampson and Morenoff 1997), that researchers began developing a theoretical framework for studying neighborhood effects. For example, Durkheim (1951 (1897)) theorizes that social forces external to the individual (e.g. norms and values) influence suicide and Wilson (1987: 56–60) links deleterious individual outcomes to the absence of middle- and working-class families in ghetto neighborhoods. Although many of the hypotheses set forth in the field's infancy are based on empirical observations, the theories are still largely abstract and too remote from particular classes of social behavior, organization and change to make strong generalizations. Earlier literature on the subject left the reader with the general impression that the structural and social characteristics of the places we live in matter, but left the theory for further development.

Given the growth in nationally representative individual-level data containing geographic indicators and other information on neighborhood characteristics, researchers aggressively set forth in estimating the effects of neighborhoods on individual well-being. A cottage industry emerged in which scholars from a variety of disciplines attempted to determine how and to what degree neighborhoods affect a wide range of individual outcomes. Many of these studies conceptualized neighborhood conditions using mean measures of socioeconomic status, such as income, poverty rate and educational attainment, while others combined these variables into a single index of neighborhood quality. These earlier studies found neighborhood characteristics to be related to a variety of individual outcomes, such as crime, health, youth academic performance and labor market success (Brooks-Gunn et al. 1997; Small and Newman 2001; Sampson et al. 2002). Based on this literature, we can reject the assertion that neighborhoods never matter.

A defining characteristic of many of these earlier studies is that the concept of space took center stage in the modeling and interpretive stages. The social sciences have a long tradition of studying phenomena at the individual level. Many factors contributed to this, including the growth of nationally representative individual-level data sets and concerns over ecological fallacy (Voss 2007). However, ecological scholars across many disciplines challenged this belief by allocating higher importance to the influence of community conditions on individual social processes (Wilson 1987: pp 165–66, Sampson and Morenoff 1997; Dorling et al. 2001). Although many measures of place in earlier neighborhood effects studies were fairly crude, they were explicitly defined in the model and took the forefront in the analysis and theory. In fact, we can trace the earlier development of many spatial concepts, procedures and statistical models, such as multilevel modeling and the spatial clustering of social processes, to neighborhood effects estimation (Voss 2007).

While these studies show that neighborhoods may matter, they do not delineate why they matter. The field began with an abstract theory about the neighborhood and ended with atheoretical findings based largely on observations and empirical results. In a survey of the literature, Jencks and Mayer (1990) bemoan the tendency of researchers to rely on a “black box” model of neighborhood effects. The field reacted by urging researchers to craft further theory on the mechanisms that link residents to their neighborhoods (Duncan et al. 1997). By understanding how neighborhoods affect individuals, researchers can tie developed theories directly to model parameters, lending stronger credence and a more nuanced understanding of significant findings.

Another force within the field was pushing research in a different direction. Researchers noted the difficulty of obtaining causal estimates of neighborhood effects because of the selection processes that confound the relationship between place and individual (Duncan and Raudenbush 1999; Sampson et al. 2002). If individuals with certain characteristics choose to live in certain neighborhoods, and those characteristics are correlated with higher or lower outcomes, then a neighborhood effect may reflect the characteristics of individuals rather than the attributes of the neighborhoods themselves.

While the encouragement for stronger theorizing on the mechanisms that link neighborhood to individuals was well received by the research community, the even stronger push to develop better research designs and statistical models to yield stronger causal inference was leading researchers further down an atheoretical road. Many researchers began searching for methods to yield stronger causal inference.

While some relied on sophisticated regression procedures like matching and instrumental variables (Foster and McLanahan 1996; Harding 2003), many considered social experiments to be the solution to many of the methodological problems earlier observational studies confronted. In particular, researchers began to interpret the treatment effects of housing mobility programs as neighborhood effects. Such a strategy administers treatment at the individual level and its effect is an example of a person-based neighborhood effect.

4.3 Person-Based Neighborhood Effects

In a person-based approach, the idea is to offer narrow programs to specific individuals within neighborhoods. The concern is to direct specific services and assistance to those highest in need. Policies such as providing education, training, job and family counseling, relocation assistance, and certain types of health care assistance form the core of person-based approaches.

The concept of person-based interventions is not new. Policymakers in the United States and abroad have developed and implemented person-based interventions for decades. One of the earlier person-based interventions in the U.S. was the Aid to Families with Dependent Children (AFDC) program, which provided financial assistance to single parents and low-income families, created by the Social Security Act in 1934. Head Start is a similar program, but targeted success at school for low-income children through nutritional, social and educational services. The defining characteristic of these programs is to provide direct assistance to those in need with little regard to any indirect effects on local communities and the general population.

While person-based interventions have a long history, their use as a strategy for estimating neighborhood effects is relatively new. Neighborhood effects researchers have primarily drawn on experimental or quasi-experimental housing assistance programs, which attempt to provide poor people living in disadvantaged neighborhoods with vouchers to move into better communities. The attractive feature of many of these interventions is their placement of individuals into treatment and control conditions typically through random assignment.

Housing vouchers were first allocated nationwide as a result of the U.S. Housing and Community Development Act of 1974. One of the first housing mobility programs in the U.S. was the Gautreaux Housing Experiment (1976) in Chicago. The program provided assistance to poor African-American families to move into census tracts with 30 % or fewer black residents. The program moved more than 7,000 low-income black families between 1976 and 1998. Researchers found that moving families improved across a variety of outcomes, including child schooling success, financial security and health well-being (DeLuca and Dayton 2009).

Building off the Gautreaux program, a slew of housing assistance programs were developed in the following years. Examples include the Thompson Baltimore Desegregation program (Engdahl 2009), the Hope VI initiative (Popkin et al. 2004) and Section 8 vouchers. These programs provided much of the same services as Gautreaux and focused on moving poor minorities to wealthier neighborhoods. Researchers evaluated the effects of many of these programs and interpreted them as neighborhood effects.

While these programs mimicked the research design of a randomized experiment by separating individuals into treatment and control groups, assignment into these groups was not explicitly random. Therefore, although the results were promising, they were viewed with skepticism since authors are unable to control for the potentially endogenous reasons why low-income households choose to apply for and eventually receive a housing voucher. The solution to these concerns is to use data collected from a randomized field experiment.

Several randomized housing assistance programs surfaced in various large, urban cities throughout the nation. Examples include the Chicago Housing Choice Voucher Program, the Yonkers Project (Fauth et al. 2007) and the Welfare to Work Housing Voucher Experiment (Wood et al. 2008), which was conducted in six U.S. cities in 2000. The largest and most expensive program is the U.S. Department of Housing and Urban Development's Moving to Opportunity (MTO) initiative, which was offered in five large cities in the 1990s. All of these programs share the common characteristic of randomly selecting families from a restricted population, usually based on income level and race, into treatment and control conditions. Unlike the neighborhood effects estimated from previous housing assistance interventions, the effects obtained from these programs were argued to be free from selection issues because of randomization.

Neighborhood researchers have interpreted housing mobility program effects as neighborhood effects (Sampson 2008). Let Y_{ij} denote the outcome of the *ith* individual living in the *jth* neighborhood, X_{ij} as individual-level characteristics for individual i in neighborhood j and, Z_j as neighborhood-level characteristics for neighborhood j. Suppose we conduct a person-based or individual-level intervention similar to the one carried out by the MTO program, where individual i is randomly selected to receive housing search assistance to move from a high to low poverty neighborhood[1] Denote the treatment variable as an indicator T_{ij} with treatment value t^* and non-treatment value t^0. For a housing mobility intervention, t^* equals 1, which indicates moving from a high to low poverty neighborhood, and t^0 equals 0, which indicates remaining in the same neighborhood. The typical model using a person-based intervention to estimate a neighborhood effect will have a multilevel hierarchy, with individual-level i given as

$$Y_{ij} = \beta_1 X_{ij} + \sigma^{ind} T_{ij} + \mu_j + \epsilon_{ij}$$

and neighborhood-level j given as

$$\mu_j = \alpha + \beta_2 Z_j + \phi_j,$$

[1] For the purposes of this example, assume all treated individuals use the assistance voucher.

where ϕ_j denotes a neighborhood-level random error and ϵ_{ij} denotes an individual-level random error, both assumed to be independent and identically distributed with mean zero.

We can use the potential outcomes framework established by Rubin (1974) to provide a causal interpretation of the individual-level treatment effect σ^{ind}. Define $Y_{ij}(t)$ as the outcome for individual i in neighborhood j if the person-level treatment assigns the individual a value of t. Since we are not concerned with the effect of treatment on individual i per se but on the population average, the treatment effect σ^{ind} equals $E\left[Y_{ij}(1)\right] - E\left[Y_{ij}(0)\right]$, which is the average outcome for those treated minus the average outcome for those not treated. Assuming T_{ij} is independent of Y_{ij} controlling for X_{ij} and Z_j and that the potential outcome for any particular individual i is stable, the treatment effect is (VanderWeele 2010):

$$E\left[Y_{ij}(1)\right] - E\left[Y_{ij}(0)\right] = \sigma^{ind}$$

Thus, σ^{ind} equals the causal effect of moving from a high to low poverty neighborhood due to a housing voucher. Many researchers and practitioners have also interpreted σ^{ind} as a causal neighborhood effect (Ludwig et al. 2008).

4.4 Place-Based Neighborhood Effects

Rather than moving individuals into better places, we can improve the neighborhoods they currently live in. Such a strategy administers treatment at the neighborhood level and its impact is an example of a place-based neighborhood effect. In a place-based approach, less narrow, far-reaching programs are introduced throughout the neighborhood with the broad intent of lifting overall neighborhood quality. Examples include business tax credits, improved infrastructure and affordable housing development. Place-based approaches largely draw on Keynesian economic theory, which supports the idea that improvements in the neighborhood benefit all residents, or that "a rising tide lifts all boats."

Similar to person-based programs, place-based initiatives are not new. The use of place-based programs in the United States go back to the mid twentieth century, when federal urban renewal programs gave cities matching funds for the removal of blighted neighborhoods and their redevelopment. Many of these programs involve some flavor of community economic reinvestment, including tax breaks for businesses, an influx of financial services in poor neighborhoods and resident employment programs. Urban enterprise zones, which offer tax incentives to attract new businesses to disadvantaged areas, and the Community Reinvestment Act (1977), which encouraged local financial institutions to help meet the credit needs of the communities in which they operate, are examples. Place-based programs have been used in a variety of other settings, including "whole-school" reforms (Cook et al. 2000) and crime prevention (Sherman and Weisburd 1995).

The most popular place-based interventions in the U.S. are community development programs, typically run by local non-profits that focus not on one target service, but on whole community redevelopment. These programs seek to tackle disadvantage across a large number of fronts, stressing the importance of employment and personal responsibility, the effects of childhood poverty and the enabling effects of strong neighborhoods and social inclusion. These place-based initiatives, particularly those focused on improving educational outcomes, have recently garnered increased attention due to the success of the Harlem Children Zone (HCZ) program in improving the overall welfare of impoverished, high-risk youths in Harlem (Dobbie et al. 2011). Borrowing from HCZ, the United States government implemented the Promise Neighborhoods initiative, which awarded planning grants to develop neighborhood wide cradle-to-career services to 21 communities throughout the U.S.

Currently, most place-based interventions are initiated by community development corporations (CDC), which have broad community betterment missions and engage in a wide variety of activities, but their signature accomplishment is the production of affordable housing. Surprisingly, there has been limited research done on the impact of CDC programs, particularly at the individual level (Vidal and Keating 2004). Galster et al. (2004) conducted three case studies of large-scale, CDC led community development initiatives in Portland, Denver, and Boston, using single-family home prices as the outcome indicator and found that in all three cases prices increased.

Neighborhood effects researchers have ignored place-based initiatives as a means for estimating neighborhood impact on individual outcomes. This may be due to the small number of place-based programs using random assignment. The Jobs-Plus Community Revitalization Initiative for Public Housing Families, which provides community-based support for obtaining and maintaining employment for individuals in public housing developments in six major U.S. cities, is one of the few examples of a randomized place-based program. Analysts selected matched candidate housing developments for each Jobs-Plus site and then randomly chose the development that would launch the intervention. The remaining developments thus serve as a comparison group. Despite being a randomized intervention, Jobs-Plus has not received any attention by the neighborhood effects literature.

We can estimate the effect of a place-based program using a multilevel model. Define T_j as the place-based intervention variable, where neighborhood j is given a value of 1 if it is treated and a value of 0 otherwise. The individual-level equation of a multilevel model for a place-based intervention is

$$Y_{ij} = \beta_1 X_{ij} + \mu_j + \epsilon_{ij}$$

and the equation for neighborhood-level j is

$$\mu_j = \alpha + \beta_2 Z_j + \sigma^{ne} T_j + \phi_j$$

The neighborhood-level treatment effect σ^{ne} under the potential outcomes framework is

$$E\left[Y_{ij}(1)\right] - E\left[Y_{ij}(0)\right] = \sigma^{ne},$$

assuming T_j is independent of Y_{ij} controlling for X_{ij} and Z_j and treatment consistency across all treated neighborhoods. The coefficient σ^{ne} measures the impact of the place-based intervention on outcome Y_{ij}. However, similar to σ^{ind} we can also treat σ^{ne} as a neighborhood effect. While the former changes the neighborhood by moving individual i, the latter changes the neighborhood by directly changing individual i's neighborhood j.

The concept of person- and place-based interventions is not new, both in academic research and in applied settings (Spencer 2004; Bloom and Riccio 2005; Verbitsky and Raudenbush 2004). However, neighborhood effects research has largely ignored the person- and place-based dichotomy, which has been articulated only recently (Sampson 2008; Ludwig et al. 2008). By clarifying how we should interpret results from experimental interventions like public housing assistance programs or community economic redevelopment initiatives, a framework for understanding and measuring neighborhood effects surfaces, specifically by how policy actually employs programs to improve individual welfare, either by person or place. The person- and place-based typology is not just a methodological concern, as it is intrinsically tied to the standard policy approaches to community- and individual-level improvement. Once we determine whether and how neighborhoods affect individuals, the ultimate goal is to understand ways to use these results to formulate appropriate social policies. The delineating feature of each approach is the target of policy investment. Person-based initiatives target individuals or households, whereas place-based ones target particular areas and neighborhoods.

Besides the placement of the treatment variable in the multilevel model, what are the differences between the neighborhood effects obtained from person- and place-based interventions? Sampson (2008) distinguishes the two approaches and emphasizes that theories aiming to explain results drawing from neighborhood-level variation is logically not the same as those explaining results relying on individual-level variation. Drawing from the current literature on mediation (Baron and Kinney 1986; Pearl 2001; VanderWeele 2010; Imai et al. 2013), the next section builds on the person- and place-based framework by decomposing the average neighborhood effect into its natural direct and indirect components. Causal mediation analysis has been used in a variety of settings, but has been largely ignored in the neighborhood effects literature despite the strong push to identify the mechanisms that connect individuals to neighborhoods.

4.5 What Do Person- and Place-Based Interventions Tell Us?

As the neighborhood effects literature continues to push for better statistical methods and stronger theory, research has gradually moved away from the standard observational analyses that were common during the field's infancy. A survey of the current literature suggests that the field is heading towards two seemingly divergent modes of research: one centered on developing more sophistical statistical and design methods to deal with the selection problems that have plagued previous studies (Harding 2003; Durlauf 2004; Sharkey 2012) and the other focused on developing theories on the mechanisms that link neighborhoods to individual well-being (Ainsworth 2002; Sampson et al. 2002; Galster 2012). Many of the new methods allow researchers to determine whether neighborhoods matter, but not why they matter. Many of the theories provide insight into how neighborhood processes may impact residents, but they have yet to be empirically tested.

Person-based interventions, specifically housing mobility programs, have received an increasing amount of attention within the neighborhood effects literature since the first results of Gautreaux and the MTO were released several years ago. Although some have cautioned against the identification of findings from these interventions as neighborhood effects (Sampson 2008; Small and Feldman 2012), the predominant view is that randomized interventions like the MTO are the field's best strategy for obtaining statistically sound evidence of neighborhood impact (Ludwig et al. 2008). However, it is still unclear what these interventions actually tell us about neighborhoods. More importantly, can we conclude that findings from person-based randomized interventions are as atheoretical as they have been currently presented? If randomizing is the best strategy for obtaining unbiased effects, what can randomized place-based interventions tell us about neighborhoods? Do they provide more insight into neighborhood mechanisms or are they similar to person-based mobility programs but packaged differently? In this section, I set up a framework that will allow us to identify how person- and place-based interventions may differ and determine what each can tell us about the mechanisms possibly underlying neighborhood effects.

The basic concern is to compare σ^{ind} and σ^{ne}. We want to decompose these average effects to understand what each treatment is telling us about the neighborhood. In order to do so, we need to decompose the effects into parts that make comparisons methodologically tractable. Fortunately, researchers have established a framework for making such a decomposition possible by formulating mediation effects of interventions (Pearl 2001; VanderWeele 2010). The concept of mediation captures the idea that interventions may (indirectly) work through various individual- and neighborhood-level mechanisms to impact individual-level outcomes.

The framework underlying the comparison is shown in Fig. 4.1. The standard approach to estimating the impact of an intervention is to measure the direct link between the intervention and individual outcomes, which yields the total average effect. Researchers commonly report this as a neighborhood effect despite it not

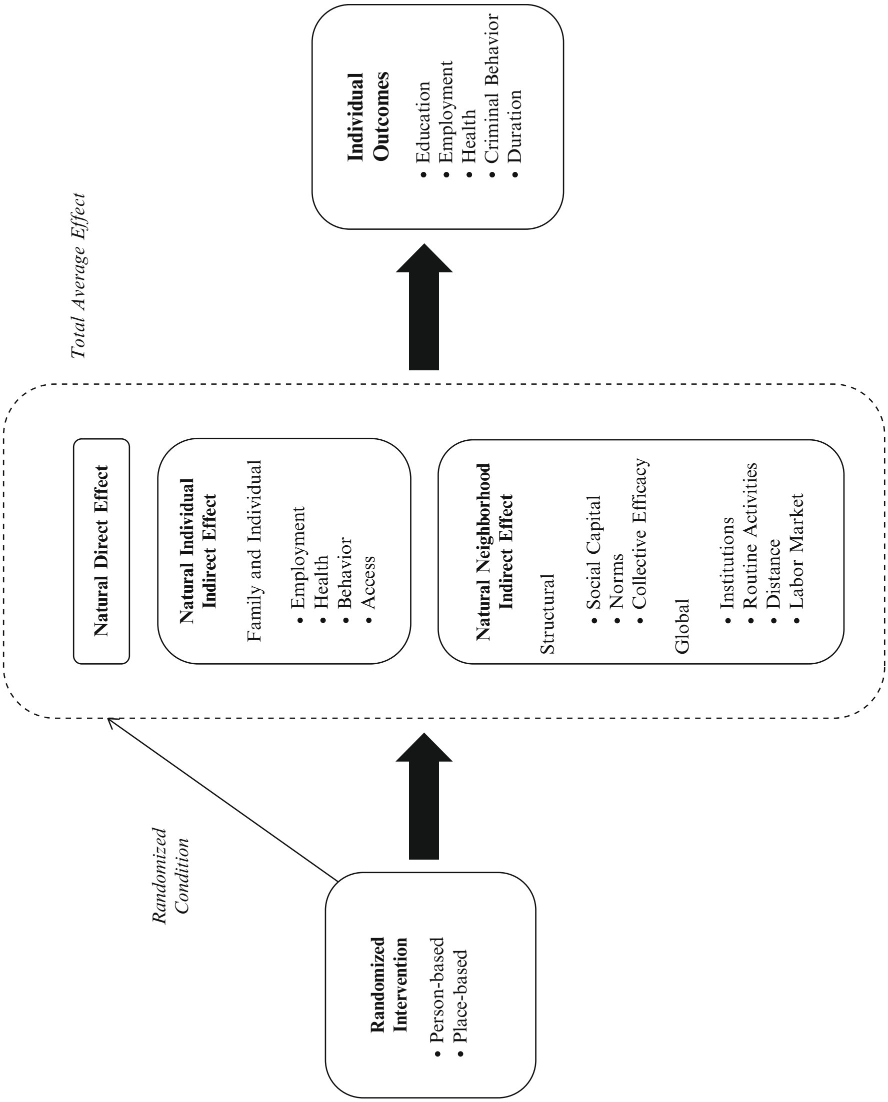

Fig. 4.1 Conceptual framework of person-and place-based intervention neighborhood effects

revealing anything specific about the neighborhood. However, the total average effect can be decomposed into several components: the natural direct effect and natural individual- and neighborhood-level indirect effects. The decomposition allows researchers to do two things: determine how and why person- and place-based neighborhood effects differ and to establish middle-range testable theories of how neighborhoods influence individual outcomes. The framework blends the two seemingly divergent research trends in the neighborhood effects literature; it allows researchers to open up the black box of interventions and see what mechanisms are activated when a person- or place-based program is enacted on a population.

For the purposes of this exposition, we want to use this framework to compare a person-based intervention similar to the MTO to a comparable place-based intervention that randomizes at the community level. In the person-based intervention, individuals are randomly selected to receive a housing assistance voucher to move into higher quality neighborhoods. In the place-based intervention, individuals are not randomly assigned to better neighborhoods, but neighborhoods themselves are randomized to better conditions.

4.6 Natural Direct Effect

Most interventions implement a set of programs to achieve goals often broad in scope. Hence, it is difficult to interpret the social mechanisms underlying their effects. An improvement in outcomes tells us that neighborhoods have an impact, but not why they have an impact. The problem is that interventions typically change an entire bundle of neighborhood characteristics, most of which were not randomized before treatment, making it difficult to disentangle simultaneous changes in structural factors and social processes (Katz et al. 2001).

The natural direct effect measures the impact of the specific condition randomly assigned by the intervention to treatment units. Any individual- and neighborhood-level characteristics subsequently or not at all randomized are mediators that affect the outcome indirectly. In housing mobility programs, individuals are randomized into two conditions: receiving and not receiving an assistance voucher to move to a lower poverty neighborhood. If we assume those who receive a voucher use it, then the natural direct effect measures the impact of moving from a poor to non-poor neighborhood or more precisely, the impact of moving using a housing voucher.

Similarly, the natural direct effect of a place-based intervention measures the impact of the randomized condition, typically some program intended to improve community quality. For example, the natural direct effect of enterprise zones measures the impact of business tax credits. Unlike for a housing mobility intervention, the natural direct effect for a place-based program makes explicit the mechanisms believed to account for the relationship between neighborhood quality and individuals. If researchers believe a strong local labor market is a neighborhood mechanism, they can explicitly test this condition through a place-based intervention by randomizing neighborhoods to a treatment that infuses job opportunities within those communities. Middle-range theories of neighborhood mechanisms like local economic conditions can be directly tested through place-based programs since the programs themselves, rather than representing an individually focused condition such as mobility, contain the actual neighborhood mechanisms that researchers want to directly test for.

4.7 Natural Individual Indirect Effect

The natural individual-level indirect effect measures the intervention's effect on the outcome mediated through individual-level mechanisms. The mediation pathway contains three steps: (1) The intervention directly changes something about the individual or neighborhood, (2) which alters the behavior or a characteristic of the individual and (3) results in a change in the outcome of interest. For example, an influx of bars and liquor stores into a neighborhood leads to higher levels of individual alcohol consumption, which then impacts cardiovascular and liver function.

We can separate the individual indirect effect into individual- and family-level components. In a person-based housing mobility intervention, moving into a lower poverty neighborhood may affect several family related characteristics that may influence an individual's well-being. For example, parents have a greater opportunity of obtaining better paying jobs in wealthier neighborhoods, thus the household income level may rise and parental health may improve. Children receive more resources and parental support and thus may perform better in school, have better health outcomes and engage in activities that deter juvenile delinquency. These examples highlight positive effects, but we can also hypothesize changes to the family that lead to negative effects, such as a reduction in parental time with children because of greater work responsibilities.

The intervention may significantly impact mechanisms connected to the individuals themselves. For example, children may engage in more physical activity due to nearer proximity to parks or enroll in academic-enhancing programs such as after school tutoring, which were not offered in previous neighborhoods. A child's exposure to physical hazards (e.g. distance to nearest highway or toxic waste plant), propensity towards crime and delinquency, and health related characteristics are other examples of individual-level variables that have an impact on a child's well-being and may be altered after changing neighborhoods.

The above examples of person-based family- and individual-level mechanisms also apply as mediators in a place-based intervention. However, there are two differences. First, person-based interventions typically target the individual, thus resources are specifically tailored to individual-level enhancement, while the effects of a place-based intervention may be spread out over all residents. Second, because place-based interventions directly work at the neighborhood level, it may take a much longer time for their effects to trickle down to individuals. When evaluating individual indirect effects we must keep in mind that they are compositional effects – researchers may regard these as separate from the true neighborhood effect since the source of variation is not due to factors intrinsic to the neighborhood but due to characteristics of individuals that happen to be clustered within certain neighborhoods.

4.8 Natural Neighborhood Indirect Effect

Neighborhood-level indirect effects are commonly referred to in the literature as neighborhood mechanisms, which are typically conceived of as contextual processes that account for how neighborhoods bring about change in a given phenomenon of interest (Sampson et al. 2002). Drawing from the theoretical work of Wilson (1987), Jencks and Mayer (1990) and Sampson et al. (1999), we can separate neighborhood indirect effects into two broad categories: structural or social-interactive mechanisms, which include social capital, norms and collective efficacy, and global or institutional mechanisms, such as institutional resources and routine activities. Social-interactive mechanisms describe how the interactions between residents and the socialization amongst community members affect outcomes. For example, norms and collective efficacy alludes to the mutual trust and shared expectations of residents. Institutional mechanisms refer to the feedback and interaction between residents and the physical features of a neighborhood. For example, routine activities factor in the impact of land use patterns and ecological distributions of daily activities on individual well-being.

Similar to individual-level indirect effects, neighborhood-level indirect effects from a place-based intervention likely require more time to surface. However, a place-based intervention may eventually yield stronger indirect neighborhood-level effects since it works directly at the neighborhood level and thus treats neighborhood mediation effects not just as externalities, as they often are in person-based interventions, but integral consequences of the programs and policies instituted within the community. Additionally, by treating the entire neighborhood rather than selected individuals, place-based interventions reduce the possibility of negative social-interactive effects. In a place-based intervention, if the treatment is successful, the positive changes in the social-interactive processes within a neighborhood affect all residents, maintaining and possibly improving neighborhood interactions.

In summary, rather than thinking of the intervention treatment effect as simply a total average effect, we can use the framework presented in Fig. 4.1 to conceptually unbundle the effect into three theoretically tractable components. Each part provides a clearer idea of what the intervention is activating and how it may or may not be related to the neighborhood. More importantly, we can tie these mediation effects to the middle-range theories that neighborhood researchers have been developing but not empirically testing. The framework melds the two current dominant modes of research in the field: the seemingly atheoretical neighborhood effects derived from randomized person-based interventions and the development of stronger theorizing around how and why neighborhoods may impact individual outcomes.

The comparison of person- and place-based programs using the above framework elucidates several important differences between the neighborhood effects estimated from the two intervention types. First, the effects of a place-based intervention likely take more time to surface at the individual level. Place-based programs are broader in scope and target a much wider population that is likely less compliant and motivated than the sample chosen for person-based interventions.

Second, it is unclear whether place-based interventions yield weaker or stronger neighborhood effects. Weaker effects may occur since place-based programs are tailored specifically for neighborhoods rather than individuals. Community-wide interventions are generally assessed through a sample of residents – not only among those sufficiently motivated to participate in the intervention. It may take time for effects to surface at the individual level and those effects may be diluted across the population. Conversely, we may expect place-based interventions to have stronger effects since they explicitly account for the relationship between individuals and their neighborhoods rather than treat the relationship as an externality.

Third, the advantage of measuring neighborhood effects through a place-based program is that researchers can explicitly test hypothesized mechanisms linking neighborhoods to individuals. In other words, the researcher can directly manipulate and assess the significance of the indirect effects represented in the second column of Fig. 4.1. In housing mobility programs, the analyst can only directly randomize a single condition: moving individuals to better neighborhoods. The individual condition is explicitly manipulated but the neighborhood condition changes only indirectly. In a place-based program, analysts explicitly enact a set of programs that they believe will enhance overall community conditions and subsequently improve resident outcomes. For example, if researchers want to test whether increased social capital within a community improves individual well-being, they can implement a place-based intervention that, for example, creates a neighborhood association that meets weekly and institutes community activities such as monthly neighborhood block parties and book clubs. If researchers want to test crime as a neighborhood mechanism, they can create neighborhood watch programs or increase police activity. A person-based program can certainly test such mechanisms, but the manipulable condition is the mobility of individuals into a community with greater social capital or less crime rather than changing community social capital or crime directly. Compared with person-based programs, place-based interventions require much greater knowledge on the part of policy makers about what specific neighborhood attributes matter most for improving outcomes.

Lastly, the juxtaposition of individual- and neighborhood-level treatments reveals the relative importance of space in the effects of both types of interventions. We can see that place-based programs explicitly model space in two ways. First, place-based interventions directly capture space by identifying communities as the policy and research design targets. Neighborhood effects derived from a person-based intervention are more likely to be contaminated since the target is the individual rather than the neighborhood. A person-based intervention that focuses on the individual without regard to his interactions with the structural and physical features of his community may not only contaminate neighborhood effect estimates, but also negatively impact an individual's well-being. Clampet-Lundquist (2007) finds that involuntary relocation severs social ties and makes it difficult for families to establish new social networks in their new communities.

Place-based programs recognize that even in a world of generous transfer payments, many low-income households are clustered in areas characterized by high levels of poverty and low levels of social, labor market and financial resources. The core assumption underlying most person-based programs is that individuals can succeed if they are given individually focused opportunities. Place-based programs come from the perspective that disadvantage can be driven by the opportunities available in the local community regardless of individual standing. Hence, the neighborhood is the policy target, and not the individual.

Second, the juxtaposition of individual- and neighborhood-level treatments reveals the potential impact of these interventions on external agents. In the case of the individual-level treatment, we estimate the effect on those who were selected to move to other neighborhoods. However, an important overlooked aspect is the effect on individuals who are not treated, i.e. those who remain in the same neighborhood. While economically beneficial to the treated individual family, this kind of move is likely to be detrimental to residents of the neighborhood from which the family moved out of. From a population-level perspective, it appears inefficient to move only selected families to better communities while leaving the old neighborhood in its original or potentially worse conditions. Although they often move, poor families remain concentrated in disadvantaged neighborhoods and thus the neighborhood effects estimated through a person-based intervention like the MTO apply to only a select group of individuals, namely families who have the opportunity to move to a wealthier community.

In a place-based treatment, since the target is the entire neighborhood, all individuals within that neighborhood are impacted by the intervention, which addresses the clustering of disadvantage Wilson and earlier neighborhood effects scholars identified in their ethnographic research. This does not necessarily translate into positive effects for all residents, as the intervention may push families out to other poor neighborhoods, but a carefully constructed neighborhood intervention has greater potential for having a wider population effect than a similarly constructed person-based intervention. The impact of an intervention is a function of its effectiveness in generating individual behavioral change, and its reach, defined as the penetration of the intervention within the population (Sorensen et al. 1998). There is likely a trade off, ceteris paribus, between the impact depth of a person-based treatment and the wider reach of a place-based intervention.

4.9 Conclusion

The early phase of neighborhood effects research posed a clear set of questions-does the neighborhood affect life chances, and, if so, how?-that provided a clear target for researchers to pursue. Empirical studies tested the first of these questions and the evidence suggests that neighborhoods matter. However, two problems emerged from this earlier work. First, results point to neighborhoods having an impact, but did not show why they have an impact. Second, earlier studies were largely observational and thus likely fall victim to selection bias.

As a response to these issues, two seemingly divergent research agendas emerged. Some neighborhood effects researchers pushed for stronger theorizing around the mechanisms that link neighborhoods to individual well-being. However, there was a need to strengthen the integrity of the results by minimizing the methodological problems that plagued earlier studies. Researchers employed a variety of statistical and research- design strategies to deal with selection bias, the most prominent being randomized housing mobility interventions. Neighborhood effects derived from such person-based programs have garnered increased attention, so much so that it appears mobility interventions may become permanent fixtures in the landscape of neighborhood effects research. However, although interventions may generate results with greater internal validity, they fail to fully specify the explicit mechanisms that connect neighborhoods to individuals. We may have a stronger answer to the question "do neighborhoods matter," but were still left pondering why they matter.

In this chapter, I outlined a framework for understanding what mobility interventions tell us about neighborhoods and their impact on individuals. While we can't use the framework to explicitly estimate the mechanisms activated by previous interventions like the MTO, it can guide researchers in setting up future housing mobility programs so that mechanisms can be directly measured and conceptualized. Additionally, the framework sheds light onto a potential alternative, place-based interventions, to estimating neighborhood effects that may provide a more suitable testing grounds for middle-range theories and minimize some of the interpretability issues that hinder person-based interventions.

The strong push in the field for estimating neighborhood effects through randomized experiments is here to stay. Moving forward, as researchers pursue new kinds of questions, they would do well to transcend the limitations of the past by developing ways to mesh theory with sophisticated statistical methods. Through this amalgam, new concepts, such as place-based neighborhood effects, emerge that will push the field towards greater theoretical cohesion and methodological integrity.

References

Ainsworth, J. W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces, 81*(1), 117–152.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Psychology, 51*, 1173–1182.

Bloom, H. S., & Riccio, J. A. (2005). Using place-based random assignment and comparative interrupted time-series analysis to evaluate the jobs-plus employment program for public housing residents. *The Annals of the American Academy of Political and Social Science, 599* (1), 19–51.

Brooks-Gunn, J., Duncan, G. J., & Maritato, N. (1997). Poor families, poor outcomes: The well-being of children and youth. In G. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 1–17). New York: Russell Sage.

Clampet-Lundquist, S. (2007). No more 'bois ball: The impact of relocation from public housing on adolescents. *Journal of Adolescent Research, 22*(3), 298–323.

Cook, T. D., Hunt, H. D., & Murphy, R. F. (2000). Comer's school development program in Chicago: A theory-based evaluation. *American Educational Research Journal, 37*, 535–597.

DeLuca, S., & Dayton, E. (2009). Switching social contexts: The effects of housing mobility and school choice programs on youth outcomes. *Annual Review of Sociology, 35*, 457–491.

Dobbie, W., Fryer, R. G., & Fryer, G., Jr. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics, 3*(3), 158–187.

Dorling, D., Smith, G., Noble, M., Wright, G., Burrows, R., Bradshaw, J., Joshi, H., Pattie, C., Mitchell, R., Green, A., & McCulloch, A. (2001). How much does place matter? *Environment and Planning A, 33*, 1335–1369.

Duncan, G., Brooks-Gunn, J., & Aber, J. L. (1997). *Neighborhood poverty: Context and consequences for children*. New York: Russell Sage.

Durkheim, E. (1951). *Suicide: A study in sociology* [1897] (J. A. Spaulding, & G. Simpson, Trans.). Glencoe: The Free Press.

Duncan, G. J., & Raudenbush, S. W. (1999). Assessing the effects of context in studies of child and youth development. *Educational Psychologist, 34*(1), 29–41.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of Regional and Urban Economics, 4*, 2173–2242.

Engdahl, L. (2009). *New homes, new neighborhoods, new schools: A progress report on the Baltimore Housing Mobility Program*. Washington, DC: Poverty & Race Research Action Council.

Fauth, R. C., Leventhal, T., & Brooks-Gunn, J. (2007). Welcome to the neighborhood? Long-term impacts of moving to low-poverty neighborhoods on poor children's and adolescents' outcomes. *Journal of Research on Adolescence, 17*(2), 249–284.

Foster, E. M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school? *Psychological Methods, 1*(3), 249–260.

Galpin, C. (1915). *The social anatomy of an agricultural community* (Research bulletin 34). Madison: Agricultural Experiment Station, University of Wisconsin.

Galster, G. (2008). Quantifying the effect of neighborhood on individuals: Challenges, alternative approaches and promising directions. *Journal of Applied Social Science Studies, 128*(1), 7–48.

Galster, G. C. (2012). The mechanism(s) of neighbourhood effects: Theory, evidence, and policy implications. In *Neighbourhood effects research: New perspectives* (pp. 23–56). Dordrecht: Springer.

Galster, G., Temkin, K., Walker, C., & Sawyer, N. (2004). Measuring the impacts of community development initiatives. *Evaluation Review, 28*(6), 1–38.

Harding, D. J. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology, 109*(3), 676–719.

Harding, D. J., Gennetian, L., Winship, C., Sanbonmatsu, L., & Kling, J. R. (2011). Unpacking neighborhood influences on education outcomes: Setting the stage for future research. In G. Duncan, & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 277–296). New York: Russell Sage; Chicago: Spencer Foundation.

Imai, K., Tingley, D., & Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society, 173*(1), 5–51.

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. E. Lynn Jr. & M. G. H. McGeary (Eds.), *Inner city poverty in the United States*. Washington, DC: National Academy Press.

Katz, L. F., Kling, J. R., & Liebman, J. B. (2001). Moving to opportunity in Boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics, 116*(2), 607–654.

Ludwig, J., Liebman, J., Kling, J., Duncan, G. J., Katz, L. F., Kessler, R. C., & Sanbonmatsu, L. (2008). What can we learn about neighborhood effects from the moving to opportunity experiment? A comment on Clampet-Lundquist and Massey. *American Journal of Sociology, 114*(1), 144–188.

Park, R., & Burgess, E. (1925). *The city: Suggestions for investigation of human behavior in the urban environment*. Chicago: University of Chicago Press.

Pearl, J. (2001). Direct and indirect effects. In J. S. Breese & D. Kohler (Eds.), *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (pp. 411–420). San Francisco: Morgan Kaufmann.

Popkin, S. J., Katz, B., Cunningham, M. K., Brown, K. D., Gustafson, J., & Turner, M. A. (2004). *A decade of HOPE VI: Research findings and policy challenges*. Washington, DC: The Urban Institute.

Raudenbush, S. W., & Sampson, R. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observations of neighborhoods. *Sociological Methodology, 29*, 1–41.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688–701.

Sampson, R. J. (2008). Moving to inequality: Neighborhood effects and experiments meet social structure. *American Journal of Sociology, 114*, 189–231.

Sampson, R. J., & Morenoff, J. (1997). Ecological perspectives on the neighborhood context of urban poverty: Past and present. In J. Brooks-Gunn, G. J. Duncan, & L. Aber (Eds.), *Neighborhood poverty: Policy implications in studying neighborhoods* (pp. 1–22). New York: Russell Sage.

Sampson, R. J., Morenoff, J. D., & Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy of children. *American Sociological Review, 64*, 633–660.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing neighborhood effects: Social processes and new directions in research. *Annual Review of Sociology, 28*(1), 443–478.

Sharkey, P. (2012). An alternative approach to addressing selection into and out of social settings neighborhood change and African American children's economic outcomes. *Sociological Methods & Research, 41*(2), 251–293.

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly, 12*(4), 625–648.

Small, M. L., & Feldman, J. (2012). Ethnographic evidence, heterogeneity, and neighbourhood effects after moving to opportunity. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Neighbourhood effects research: New perspectives* (pp. 57–77). Dordrecht: Springer.

Small, M. L. & Newman, K. (2001). Urban poverty after the truly disadvantaged: The rediscovery of family. *The Neighborhood, and Culture, 27,* 23–45.

Sorensen, G., Emmons, K., Hunt, M. K., & Johnson, D. (1998). Implications of the results of community intervention trials. *Annual Review of Public Health, 19*, 379–416.

Spencer, J. H. (2004). People, places and policy: A politically-relevant framework for labor market efforts to reduce concentrated poverty and joblessness. *Policy Studies Journal, 32*(4), 545–568.

VanderWeele, T. J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods and Research, 38*, 515–544.

Verbitsky, N., & Raudenbush, S. W. (2004). *Causal inference in spatial setting*. Proceedings of the American Statistical Association, Social Statistics Section (pp. 2369–2374), Alexandria: American Statistical Association.

Vidal, A., & Keating, D. (2004). Community development: Current issues and emerging challenges. *Journal of Urban Affairs, 26*(2), 125–137.

Voss, P. R. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*(5), 457–476.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago.

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review, 76*(5), 713–736.

Wood, M., Turnham, J., & Mills, G. (2008). Housing affordability and family well-being: Results from the housing voucher evaluation. *Housing Policy Debate, 18*(2), 367–412.

Chapter 5
From Aspatial to Spatial, from Global to Local and Individual: Are We on the Right Track to Spatialize Segregation Measures?

David W. Wong

5.1 Introduction

Formal development of segregation measures is often associated with the classic piece by Duncan and Duncan (1955), which introduced the dissimilarity index *D* that is easy to calculate, but encapsulates insightful meanings, such as the proportion of population that needs to be reallocated in order to achieve no segregation. Some may dispute if the *D* index by the Duncans was the first segregation measure proposed, as Bell (1954) proposed another index shortly before them. Jahn et al. (1947) suggested four criteria for measuring segregation and the last criterion was used by the Duncans to build their famous *D* index. However, few will disagree that the *D* index and together with the availability of census data at that time launched the measurement venture in segregation studies. The literature is clear that the history of measuring segregation has been heavily dominated by sociologists-demographers at the early stages. The intermittent contributions by geographers in the 1980s were recognizable, but after 1990, geographers seemed to establish a niche in this venture.

Throughout these decades, scholars from multiple disciplines have been involved in "perfecting" segregation measures. Early stage of research investigated various properties of segregation measures (e.g., Cortese et al. 1976; Taeuber and Taeuber 1976; Winship 1978). These discussions on index properties were useful in affirming the versatility of the *D* index. However, White (1983) might be given the credit of using a highly hypothetical but effective checker-board landscape to challenge the robustness of the *D* index, or more accurately, all segregation indices without location or spatial information, revealing the aspatial nature of most segregation measures at that time. White and other geographers were the earliest

D.W. Wong (✉)
Geography and GeoInformation Science, George Mason University, Fairfax, VA, USA
e-mail: dwong2@gmu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_5

in introducing segregation measures incorporating spatial information explicitly (Jakubs 1981; Morgan 1983a, b), followed by a flurry of activity in the 1990s in developing spatial measures of segregation. Apparently, missing the spatial dimension in the traditional measures triggered all these individual and systematic efforts in "spatializing" segregation measurements.

The title of this book volume is "Recapturing Space". Space or geography was clearly not explicitly considered in the early formulations of segregation measures, and the effort in spatializing segregation measures has been a reaction to such a methodological deficiency. However, spatializing indices is viewed by some to be still not sufficiently spatial in the business of measuring segregation (e.g., Johnston et al. 2009). Such a call for "putting some more geography in" is partly because the spatial-geographical dimension should be clearly included in the multi-facet nature of segregation. The notion of segregation has been interrogated from the operational-measurement and conceptual perspectives. For instance, earlier discussions attempted to determine the conditions for no segregation and perfect segregation in respect to different segregation measures, particularly if random distributions of population groups are sufficiently to be labeled as no segregation (e.g., Reiner 1972; Taeuber and Taeuber 1965; Winship 1978). On the other hand, the review by Massey and Denton (1988) that drills on the conceptual meaning of segregation may be regarded as a milestone piece, which examined broadly empirical studies of segregation and deduced the five dimensions of segregation (evenness, exposure-isolation, centralization, concentration and clustering). Although results of the follow-up study were slightly more ambiguous than the original one (Massey et al. 1996), overall, the five dimensions and the position of D were ascertained. Subsequent debates joined by geographers argued if the five dimensions are distinct while the consensus that the clustering dimension is a spatial dimension seemed to have emerged. In the geographical science or spatial statistics arenas, clustering is often associated with spatial autocorrelation. Thus, a trend is the increasing use of spatial autocorrelation measures to indicate segregation. So, *does it mean that segregation is the same as having high positive spatial autocorrelation?*

Despite the multidisciplinary effort in understanding the nature of segregation and in developing effective segregation measures, we are still in the midst of failing to precisely describe conceptually what segregation is or to provide operational definitions of segregation with clarity. While the term has been used by different people, under different contexts, and for different purposes, does segregation has something in common across all these situations? Apparently, segregation may mean positive to someone, but negative to others (Peach 1996). A common notion of segregation perceived by the public and that has appeared in the literature is the separation of population groups (Newby 1982). One may perceive separation as intrinsically spatial, but space can be in different domains (e.g., White et al. 2005). If two population groups are separated within a small room, they have to be allocated to different sections of the same room, spatially apart from each other. However, separation in other contexts may not be explicitly geographical. To what extent existing measures, spatial or aspatial, are effective in capturing the separation between groups?

This chapter is intended to serve several objectives. Despite the many spatial measures that have been introduced, approaches to spatialize segregation measures are limited to just a few. Therefore, an objective of this chapter (in the next section) is to reflect systemically upon how space was (re)introduced into measuring segregation from multiple fronts. While these proposed measures are spatial in nature, it is not clear if these measures or approaches to spatialize segregation measures are sufficient to capture the basic notions of segregation, and the spatial dimension of segregation. Particularly, spatial autocorrelation or association statistics are often used in identifying spatial clusters and thus in measuring segregation. The practice of cluster determination using these spatial measures and their relationships with segregation will be critically reviewed in the third section.

Through these critical reviews, we may get closer to a "definition" of segregation and to identify the critical elements that form the foundation of measuring spatial segregation. Such elements will be enumerated in Sect. 5.4. Such critical elements can be used to develop meaningful spatial segregation measures in the future or to evaluate if a particular spatial measure is effective in capturing the essence of segregation. Thus, using these fundamental characteristics, Sect. 5.4 of this chapter also provides an assessment of the state of measuring spatial segregation. However, measuring segregation is a highly data-dependent venture. The history shows us that the early development of segregation measures was partly driven by the availability of census data. Types of data available surely facilitate and constraint how we may measure segregation. Section 5.4 will also speculate the directions of measuring segregation in the light of future data availability, the likely trend of less relying on aggregated or census-type data, and the increasing use of individual-level and survey-based data.

5.2 Geographic Space in Measuring Segregation

While the desirability of the *D* index was not unchallenged, using the index as the gold standard in segregation study was affirmed by Massey and Denton (1988) and subsequent studies. Their claim that the *D* index is the best to capture evenness, the most important dimension of segregation, was particularly influential. The *D* index is likely the most frequently used measure of segregation, but the review by Reardon and Firebaugh (2002) focusing on the distributional properties of segregation measurements endorsed the entropy-based diversity index. Following possibly the earliest idea of modeling segregation using an interaction approach by Bell (1954), Lieberson (1981) proposed the exposure-isolation indices.

Interspersed between the introductions of these measures, reviews and endorsements, several studies have warned the limitations of the *D* index in depicting segregation from a spatial perspective. These warnings, to a large degree, are also applicable to most segregation measures proposed by sociologists-demographers.

The aspatial nature of *D* and similar indices was likely first recognized by Reiner (1972) from a point pattern analysis perspective. However, one of the strongest demonstrations about the deficiency of the *D* index was the checkerboard landscape suggested by White (1983). In such a landscape, alternate areal units are exclusively occupied by one of the two population groups. Regardless of how the areal units are spatially rearranged, with neighboring units belonging to the same group or different groups, as long as each areal unit is occupied exclusively by one group, the *D* index indicates perfect segregation (=1). The checkerboard concept was exemplified by Morrill (1991) and Wong (1993) to illustrate the needs of spatial measures. The ineffectiveness of *D* and most segregation measures in reflecting the spatial separation of population groups can be boiled down to two fundamental but related spatial issues: (1) most segregation measures treat the boundaries of areal units as absolute barriers between areal units such that people on two sides of the boundaries are completely separated, despite many of those boundaries are artificial and created for the purposes of statistical enumerations such as the Census; and (2) most segregation measures do not include any spatial information depicting the spatial relationships between population groups or areal units. People residing in neighboring units are treated no differently from those on the other side of the city or region when these populations are evaluated using traditional segregation measures.

The measurement of a social phenomenon is dependent upon the data capturing the phenomenon. However, data are collected based upon certain models of the society, representing the reality in an abstract manner. Census data have been supporting the development of most segregation measures, and these data have been gathered using a model that space can be partitioned into areal units or polygons within which are the populations of concern. The census model for tabulating population statistics partly creates the impermeable boundary problem discussed above. The model, which is in fact a typical model used by geographers to organize space, also creates a challenging problem of failing to obtain consistent results in measuring segregation when population counts can be tabulated for areal units of different sizes for the same region.

Population data are provided for various types of areal units, cutting across all levels of the census geography of the U.S. and across various types of administrative, political and statistical units. However, using data tabulated for these different types of areal units to evaluate segregation will yield different results. This problem falls into the broad umbrella of the modifiable areal unit problem (MAUP), which involves the two sub-issues of scaling and zoning problems (Wong 2009). The scaling problem in segregation refers to the inconsistent segregation values obtained from using data tabulated for areal units of different sizes, such as census tracts versus census block groups. Zoning problem refers to the inconsistent results due to the use of different zonal partitioning systems, such as health service areas (HSAs) versus public use microdata areas (PUMAs), but with units of comparable sizes. In general, segregation levels are higher when smaller areal units are used, and vice versa (Wong 1997). Although a relatively thorough discussion of the MAUP on segregation did not appear until 1997 (Wong 1997) and the term MAUP was not coined until 1979

(Openshaw and Taylor 1979), Reiner (1972) already pointed out the "variability of the index with size of reporting unit" (p. 145). While the MAUP for segregation measurement is spatial in nature, one may expect that spatial segregation measures may offer some promises in producing more consistent results. Unfortunately, existing studies have not provided a more optimistic outlook in terms of using spatial segregation measures to mitigate the MAUP effects for segregation (Wong 2004). Nevertheless, future research should explore the potentials of using spatial measures to overcome the MAUP effects for segregation.

The literature is very rich in developing segregation measures and these measures are almost with no exception are indices, aspatial and spatial. While Johnston et al. (2009) called for putting more geography in measuring segregation, they are critical of the utilities of using indices in describing segregation levels. They argued that indices beyond the simple ones, such as the dissimilarity index *D*, are often too complicated to be interpreted meaningfully, and also these indices fail to reflect the segregation patterns beyond just a number. They favored the richness of group relationships that are captured by the concentration profile, which is a plot of the percent of population found in areas with increasing threshold levels (in percent) of a given population group (Poulsen et al. 2002). Recognizing the aspatial nature of the concentration profile as a major weakness of the non-index approach to measure segregation, they suggested including spatial autocorrelation statistics to capture the spatial clustering of population distribution, complementing the information derived from the profiles. While their major arguments against the use of indices sounds reasonable and have certain grains of truth, their solution of combining the profile and spatial autocorrelation statistics is equally problematic. The misgivings of using spatial autocorrelation statistics in segregation studies will be elaborated in detail in a later section.

5.3 Approaches in Spatializing Segregation Measures

The need to include (more) space or geography in measuring segregation is quite apparent. Despite the limitations of indices in reflecting spatial segregation, using indices to measure segregation is still the dominant method in segregation studies. Therefore, the discussions below will focus on indices, putting aside other methods such as graphical representations. The literature shows a variety of efforts in incorporating space in segregation indices. However, these methods may be grouped into several conceptual types.

5.3.1 Interaction Approach

Due to the dominance of the dissimilarity index *D* in segregation studies, many efforts in spatializing segregation indices, especially some of the earlier ones, revolved around the D index. A particular approach is to introduce some spatial components into the original D index formulation. Intrigued by the checkerboard pattern suggested by White (1983) to some extent, Morrill (1991) pointed out the D index (in fact, also true for most aspatial indices introduced so far) considers only population mixes within areal units, but not among areal units. Implicitly, these indices treat enumeration unit boundaries, based upon which population data are tabulated and reported, as absolute barriers such that population groups across units boundaries cannot interact. While some unit boundaries, such as rivers or political boundaries are significant barriers to impede people crossing over, boundaries of most statistical units are in fact different types of streets that are not absolute barriers for cross-over. The damning role of unit boundaries in prohibiting interaction across units is critical in limiting the effectiveness of most indices in measuring spatial segregation. The notion that people cannot cross boundaries to interact serves as the major impediment which many spatial segregation measures attempt to remove.

To overcome the artificial barriers imposed by unit boundaries, two general methods have been proposed in the literature to account for the possible interaction of different population groups across unit boundaries. The first approach is to introduce additional components to existing aspatial indices to reflect factors that may affect the potential interaction across units. The second approach is to manipulate the population data such that the population counts of areal units include to a certain extent populations in the neighboring units. Thus, interaction of populations across units is implicitly accounted for by the population data.

Using the first approach, Morrill (1991) proposed to consider the difference between the group mixes in neighboring units as the magnitude of potential interaction that can moderate the segregation as if people in neighboring units may cross over to interact. Thus, the D index was modified by an additional term reflecting the differences in racial-ethnic mixes between neighboring units. Picking up on the general notion of facilitating populations in neighboring units to interact, Wong (1993) suggested that the interaction across units may be affected by the length of the shared boundaries and the shapes of the neighborhoods. While these factors may be relevant, these extensions of Morrill's framework are rather trivial. In fact, these implementations of enabling interaction between neighboring units are quite clumsy and inefficient.

The second approach to account for the interaction of populations across units can be represented by the composite population concept introduced by Wong (1998). The general idea is to "internalize" the populations in neighboring units. To implicitly account for the potential interaction between a reference unit and its neighboring units, the population count of the reference unit should be modified to include populations in the neighboring units as if populations across the unit boundaries are freely interacting with each other. Thus, the composite population count of areal unit i, CP_i, is the sum of population in unit i and the populations in all neighbors. Formally, $CP_i = \Sigma k(P_j)$ where $k(.)$ is a neighborhood function, defining how the population in the respective unit j is counted toward the composite population of areal unit i. Note that j usually includes i itself. Using CP_i, traditional measures, such as the dissimilarity index D, exposure measures, and the diversity index can be computed, but these measures now implicitly account for population in neighboring units and therefore these index values are spatial in nature. Adopting the composite population count concept but using a binary neighborhood function of the first order (i.e., populations in the first order neighbors are counted toward the composite population of the reference unit), Wong (1998) developed the spatial version of D for multiple population groups, modified the aspatial multi-group D suggested by Morgan (1975) and Sakoda (1981).

The composite population count concept has been modified in the formulations of several spatial measures for segregation. Instead of using a binary neighborhood function of the first order, Reardon and O'Sullivan (2004) suggested using a spatial proximity function at the reference unit to determine the weights for which populations in neighboring units are counted toward the reference unit. Thus, a family of spatial version of traditional segregation measures was introduced. Conceptually, using the spatial proximity function to weight neighborhood populations is the same as the composite population count notion of removing the artificial boundaries between units to facilitate cross-unit interaction. The proximity function approach is more elegant in specifying to what extent neighboring populations should be included. Rather than "in or out" in the original CP_i formulation, the proximity function provides weights such that larger weights are for closer neighbors and smaller weights are for farther neighbors.

The composite population count or the proximity function concepts were also adopted in developing a series of spatial global and local segregation measures (Feitosa et al. 2007). Implementing these concepts removes the restrictions of boundaries in prohibiting interaction across units, implicitly accounting for the population compositions of neighboring units in evaluating segregation. This is a generic approach to spatialize the evaluation of segregation by manipulating the data. The data thus can be used for calculating any segregation measure, and the results are spatial measurements even though the formulations of those measures may not have any spatial component. Note that the exposure index by Lieberson (1981) measures the interaction between population groups within, but not across areal units.

5.3.2 Distance Approach

Modeling the interaction between population groups across units either explicitly or implicitly put space into segregation measurement. Another spatial aspect of segregation is the separation between population groups. Distance is an obvious measure to reflect the spatial separation between people. The literature has shown that distance can be included in measuring segregation at least in two approaches: distance is incorporated into existing aspatial measures and spatial measures that are formulated based upon distance.

A highly desirable property of the D index is that it indicates the proportion of population that needs to be moved in order to achieve no segregation (which definition is subject to debate). Jakubs (1981) pointed out that the effort of achieving no segregation as reflected by D does not take into account the distance involved in relocating population. Thus, he proposed a modified D (distance-based segregation index – *DBI*) by incorporating a distance measure, the total distance that the two population groups have to migrate to achieve no segregation according to the formulation of D. Soon after, Morgan (1983a) proposed another distance-based index, the modified distance-based index (*MDBI*), following the same logic used in by Jakubs, but with a different definition of complete segregation.

As mentioned earlier that Lieberson's exposure or interaction index evaluates interaction between groups within areal units, and therefore is aspatial in nature. Morgan (1983b) spatialized the Lieberson index by introducing a term reflecting the contact rate C_{ij} for populations between two areal units i and j. The contact rate can be derived based upon surveys or empirical studies. Else, based upon spatial modeling literature, the contact rate can be determined by a spatial decay model in which interaction level decreases as distance increases, but the rate of change may vary, controlling by a distance decay parameter.

While the *DBI* and two measures proposed by Morgan (the *MDBI* and the distance-decay interaction index) incorporate distance to modify existing aspatial measures, distance may be used as the basis to formulate spatial segregation measures. Recognizing the limitation of aspatial measures in handling the checkerboard-type problems, White (1983) proposed a proximity index, which is the ratio between the average intra-group separation and the average separation of all people in the study region, where the separation between people is captured by distance.

In a more general context of evaluating spatial patterns, Wong (2011) proposed a framework that encompasses spatial proximity and spatial autocorrelation, two aspects of a spatial pattern. Mirroring the typical approach to measure spatial autocorrelation that a spatial weights matrix (W) is used to control how attribute values in units are compared, he suggested that an attribute weights matrix (M) can be used to select observations meeting the attribute thresholds to be evaluated for the spatial proximity of areal units. In the M matrix, an element m_{ij} reflects the weight for the attribute similarity between units i and j. These weights can be multiplied by distance between the two units to provide the proximity value, reflecting how close those units are spatially if the attribute values of units have a certain level of similarity. The proximity value is part of the proposed *MW* index. While the M matrix can generally be applied to interval-ratio data, it can be used for ordinal and nominal data. Wong (2014) demonstrated that this framework can be used to evaluate the spatial proximity between population groups by selecting specific pairs of population groups for evaluation. The specific formulation of this general spatial pattern measure is different from White's proximity index, but its similarity to the proximity index in using the underlying notion of distance to reflect spatial segregation between population groups is apparent.

One may argue that the formulation of spatial segregation measures proposed by Reardon and O'Sullivan (2004) was based upon the concept of density and therefore grouping their work under the "interaction-based" approach seems inappropriate. Such argument is valid to a certain extent as the actual segregation measures suggested by Reardon and O'Sullivan were essentially density measures. However, to spatialize their measures, they used the proximity function (based upon distance), which enables the inclusion of neighboring population in evaluating segregation. On the other hand, the density approach was adopted by O'Sullivan and Wong (2007) in generating surfaces for different population groups so that their relatively concentration or density levels can be compared in different locations. Such approach is implicitly spatial and has the advantage of treating space as continuous.

5.4 Spatial Scale and Zonal Dependencies in Measuring Segregation

The above discussions focus on how space can be introduced into segregation measurement by either modifying existing aspatial indices or formulating spatial indices. However, capturing space is not the only spatial aspect in measuring segregation. As for most quantitative measurements in spatial sciences, measuring segregation is scale-dependent and zone-dependent, which together constitute the MAUP discussed briefly above. The literature has addressed the MAUP effects on measuring segregation quite thoroughly (e.g., Krupka 2007; Shuttleworth et al. 2010; Taylor et al. 2010; Wong 1997; Wong et al. 1999), and solutions have been sought, such as using spatial measures (Wong 2004) and decomposing segregation values to multiple scale levels (Wang 2012; Wong 2003). While the inconsistencies in segregation levels across scales and zonal configurations create tremendous challenges and inconvenience in empirical studies and analyses, they also offer opportunities to exploit these spatial aspects of measuring segregation. Several directions of segregation studies fall into this line of inquiries.

5.4.1 Global vs. Local vs. Individual

Almost without exception that early studies of segregation were intended to compare the segregation levels between cities or metropolitan areas, using census tract data as the basic units of calculation. Thus, most asptial segregation measures introduced by sociologists-demographers served the purpose well, as they are summary measures using one value to capture the level of segregation for the entire study area. During the 1990s, a movement in quantitative geography was to depict and analyze local spatial patterns (e.g., Fotheringham 1997), stemming from a series of research introducing local spatial autocorrelation statistics, statistics that reflect the extent that values in local neighborhoods are similar to each other (Getis and Ord 1992; Anselin 1995). This direction of spatial statistics led to the popular practice of detecting local clusters or hot-spot analysis.

Following this general direction of developing local statistics, Wong (2002) suggested labeling those traditional aspatial measures as global measures, parallel to those global spatial autocorrelation statistics such as Moran's *I* and Geary Ratio, and called for the formulation of local segregation measures. Based upon the concept of exposure-interaction captured in Lieberson's exposure index (which is aspatial) and Morgan's distance decay interaction index (Lieberson 1981; Morgan 1983b), Wong (2002) proposed a set of spatial local segregation indices. While these indices have been used as covariates to explain several types of health outcomes, such as low birth rates (Grady 2006) and hypertensions (White et al. 2011), these indices have some undesirable distribution properties (e.g., highly skewed and sensitive to the size of the study region). Feitosa et al. (2007) also developed a family of local segregation measures, parallel to the popular aspatial measures of dissimilarity, exposure and diversity. To develop the local spatial measures, they decomposed the global measures using approaches similar to that proposed by Wong (1996). To spatialize the local measures, they used the population intensity notion, the concept very much the same as the composite population count suggested by Wong (1998) and the proximity function suggested by Reardon and O'Sullivan (2004).

Traditional studies in segregation rely heavily on ecological or spatially aggregated data, especially census-type data, and thus analysis results are based upon data reflecting the characteristics of an area, such as a neighborhood, not individual experience. In addition, such studies relying on ecological data often focused on the population characteristics in the residential space. Segregation in the residential space definitely has significant implications and ramifications on other social and economic dimensions. Recent segregation studies expanded the scope to be more comprehensive by including various socio-geographical spaces that individuals may experience. Such call was also supported by the argument that individuals residing in the same neighborhood do not necessarily experience the same levels of segregation if we take into account of their other socio-geographical spaces, such as work space and cultural-entertainment space, avoiding some aspects of committing ecological fallacy. Using such approach and borrowing ideas from activity space research, Wong and Shaw (2011) demonstrated how individual-level data of travel patterns can be used to evaluate segregation. The segregation measure adopted was a version of the interaction index, but significantly modified to accommodate the individual nature of the data. Farber et al. (2012) also followed the local-individual approach, but relying more on the spatial autocorrelation statistics in measuring linguistic exposure between individuals of different population groups. Similarly, using a spatial association measure for categorical variables such as population groups, Páez et al. (2014) evaluated the segregation at the individual level based upon historical census records for three selected U.S. cities. All these studies illustrate that segregation level can be evaluated at the individual level. An interesting property shared by the proximity index suggested by White (1983) and one form of the *MW* index reflecting spatial proximity by Wong (2011) is that both types of measure can be used for aggregated and individual-level population data.

5.4.2 Scale-Dependent Segregation Analysis

The above discussion reports the shift from computing segregation measures for large regions such as cities, to local neighborhoods or census units, and then to individual levels. Spatially, it is a trend of zooming in to finer scales, all the way to individuals. However, segregation analysis is scale-dependent. Aggregating results from individuals do not necessarily match the results using data at the neighborhood level, and same is true between the results from the neighborhood (local) level and regional (global) levels. To a large degree, segregation level varies according to the definitions of neighborhoods. The traditional aspatial measures adopt the neighborhood definitions provided by the census geography, using census enumeration-statistical units such as census tracts or block groups. When researchers started spatializing aspatial segregation measures (e.g., Wong 1998; Reardon and O'Sullivan 2004), they redefined neighborhoods by expanding them beyond the original/reference units. Apparently, neighborhoods can be defined according to different sizes, and therefore, the segregation results are also different. Or if neighborhoods of multiple sizes are analyzed, we may realize how segregation levels may change over different spatial scales.

Accepting the fact that segregation varies with scales or sizes of areal unit, Wong (2005) suggested using neighborhoods of different sizes, from the first order neighbors to higher order neighbors in computing spatial segregation measures. As expected, the larger the sizes of neighborhoods, the more heterogeneous are the population mixes and therefore the lower the levels of segregation. While this general trend is expected, such depictions also provide insights on how segregation levels vary by spatial scales. If population groups cluster extensively, segregation levels may not decline until the sizes of neighborhood become very large, large enough to include populations of different groups outside of the clusters. Since then, quite a few studies adopted this approach by varying the sizes of neighborhoods to evaluate the variation of segregation levels over spatial scales (e.g., Reardon et al. 2008, 2009; Poulsen et al. 2010). Recently, a software package, Equipop, was developed and used in analyzing segregation at multiple spatial scales (Östh et al. 2014). The package allows users to aggregate population to neighborhoods of specific sizes before computing segregation measures, providing a convenient and flexible mechanism to aggregate areal units to neighborhoods of different spatial scales. Segregation levels at different geographical scales can then be computed and compared with some levels of control on neighborhood definitions.

5.4.3 A Brief Summary

The term segregation is commonly perceived to possess some spatial characteristics of the population, particularly about how different groups of people are spatially separated. Traditional studies of segregation use aggregated data tabulated for census areal units (such as tracts) to compute segregation indices for the entire city for cross-city or region comparisons. The aspatial nature of many segregation studies is mainly attributable to the fact that areal unit boundaries are implicitly treated as absolute barriers such that populations across units are expected not to be mixed or have interaction. Thus, segregation measures do not consider the population mixes in neighboring units. Space was introduced to measuring segregation by incorporating spatial elements into existing aspatial measures or developing new spatial measures. Population mixes in neighboring units can also be considered in evaluating segregation by creating population counts that include neighboring populations. Spatial scale also plays an important role in measuring segregation, as segregation can be measured at multiple scale levels: global, local and individual. In addition, scale is treated explicitly in measuring segregation as we can link segregation levels of a region to the scales of neighborhood based upon which segregation is evaluated.

While many different spatial elements can be introduced spatialize segregation, they are nonetheless limited to a few spatial properties: adjacency or first order neighbors, higher order neighbors, and distances between areal units, where the locations of units can be defined in various manners. All these spatial properties have been used in specifying the spatial weights matrices in spatial statistics (e.g., Bavaud 1998; Griffith 1996). The spatial weights matrices have been used to specify the spatial relationships between geographical features so that correlations between values can be evaluated in respect to their locations. In other words, to spatialize segregation measures, all we need is to include some elements capturing the spatial relationship between units or observations in the segregation measures.

5.5 Dimensions of Segregation – A Revisit

Apparently, some segregation studies using spatial measures that do not fit into the above categorization very well, particularly, those studies using spatial autocorrelation measures to focus on the clustering dimension of segregation. In the following section, let us revisit the dimensions of segregation with particular focus on the spatial aspects of segregation.

To a large degree, the five dimensions of segregation proposed by Massey and Denton (1988) have structured the majority of methodological inquiries in segregation in the past several decades. Despite this issue was revisited and previous findings were reaffirmed with some degree of fuzziness (Massey et al. 1996), the claim that segregation composes of five distinct dimensions of evenness, exposure-isolation, centralization, concentration and clustering has been a subject of debate among geographers. Johnston et al. (2007) challenged the validity of adopting the five-dimension framework and attempted to validate the presence of five dimensions empirically using U.S. data in three censuses (1980, 1990, and 2000). They claimed that evenness was identifiable, but clustering and isolation co-vary most of the time and so do concentration and centralization. Eventually, they argued that evenness, clustering and isolation share significant overlap and therefore can form the super-dimension of separateness, while concentration and centralization may form another superdimension of location.

Besides the effort by Johnston and his collaborators in empirically validating the presence of the five distinct segregation dimensions, a few conceptual discussions came to the conclusion that the five dimensions are not distinct, but can be collapsed into fewer (composite) dimensions. Reardon and O'Sullivan (2004) argued that the five dimensions can be combined to form the two conceptual dimensions of spatial isolation-exposure and spatial (un)evenness, which includes clustering, concentration and centralization. On the other hand, Brown and Chung (2006) argued conceptually that the five dimensions can be collapsed into two composite dimensions of concentration-evenness and clustering-exposure, leaving out centralization as it is no longer relevant these days as minorities are not concentrated in the central cities of large metropolitan areas given the typical polycentric city structure.

Among these discussions on the "true" dimensions of segregation, several points are worth-mentioning. First, note that some of these later discussions (e.g., Brown and Chung, 2006) about the dimensions of segregation did not recognize or acknowledge arguments made in earlier discussions (e.g., Reardon and O'Sullivan 2004) about how dimensions can be collapsed. Thus, the discussions over time were not quite coherent. Second, while the dimension labels originally used by Massey and Denton have some spatial connotations, measures representing those dimensions, with the exceptions of the measures for the centralization and clustering dimensions, are aspatial in nature. In other words, regardless how spatial those dimension labels may be, their associated measures do not capture the spatial distributions of population effectively. Although Reardon and O'Sullivan (2004) attempted to "spatialize" some of these labels by adding "spatial" in front of "evenness", "exposure" and "isolation" (p. 125), the actual measures did not become more spatial, and thus these discussions are likely not too fruitful. Clustering may be regarded as unambiguously spatial, but its meaning and definition are not as clear as most people expect, even though the term has been used frequently in segregation studies and spatial analysis. Third, Johnston et al. (2007, p. 500) claimed that their separateness and location dimensions matched the spatial exposure and spatial evenness dimensions suggested by Reardon and O'Sullivan (2004), but in fact, the underlying structures of the two sets of composite dimensions in terms of the basic five dimensions do not match. *In other words, we probably are more confused than before about the real dimensions of segregation.*

In the original analysis conducted by Massey and Denton (1988), its sequel (Massey et al. 1996), and the reanalysis (Johnston et al. 2007), a set of spatial measures, including White's proximity index, was used to reflect the clustering dimension. However, in spatial science, measuring clustering seems to be dominated by the use of spatial autocorrelation statistics, both the global and local versions. Particularly, the local versions (e.g., Getis's G and Anselin's LISA) have been used extensively to detect local clusters, and subsequently in measuring segregation (e.g., Brown and Chung 2006; Poulsen et al. 2010). Along this trend of using spatial autocorrelation statistics to measure segregation, two points need to be raised. First, spatial autocorrelation refers to the similarity levels of near by values. Highly similar values close by give strong positive spatial autocorrelation, and very dissimilarity values near each other give strong negative spatial autocorrelation. In general, a trend surface distribution produces very high positive spatial autocorrelation (Wong 2011). Do such patterns really give us the highest spatial segregation? Lee and Culhane (1998) provided an interesting argument against such perception. Second, several studies employed local spatial autocorrelation measures to identify spatial clusters of specific groups. In general, taking the general methodology of comparing one group with all other groups, percent of one group is often used in identifying the local clusters (e.g., Brown and Chung, 2006; Poulsen et al. 2010). If the percent of a specific group forms a local cluster, can we claim that the area has high segregation level? The population of that specific group may be highly concentrated in certain areas forming clusters, however, we do not know if any other group may also have high concentrations in the same areas. For instance, areas of high diversity may have high concentrations of multiple groups, and the local spatial autocorrelation levels of each group can be high in the same locations. Thus, the presence of local clusters may not necessarily reflect high segregation levels.

The literature seems to move toward a consensus that clustering is a rather distinct and spatial dimension even though the concept may be applicable to a single group. On the other hand, the exposure-isolation dimension clearly involves more than one group. Thus, these two dimensions are quite different. However, if we examine the formulation of the dissimilarity index $D = 0.5 * \sum \left| \frac{a_i}{A} - \frac{b_i}{B} \right|$, the ratios, a_i/A and b_i/B, essentially reflect how the respective groups are concentrated in unit i. Thus, |.| evaluates the difference between the concentrations of the two groups in each unit. Large absolute differences mean that one group is disproportionally represented in the corresponding units. In Massey and Denton's terminology, these disproportions create unevenness. However, if units with large disproportions are dispersed over the study region, AND if the populations can interact across unit boundaries as if there is no physical barriers to prohibit inter-unit interaction, then such situation should not be alarming, and should not constitute a highly segregated situation. Only if units with disproportions are close together, or cluster, then interaction across units will not reduce inter-group separation, and thus the situation will likely create some moderate levels of segregation. In other words, aspatial (un)evenness is based upon differences in concentration levels, and evenness by itself does not necessarily constitute segregation, but the clustering of units or areas of high concentration relative to another population group creates segregation.

5.6 Challenges and Promises in Measuring Spatial Segregation

5.6.1 Challenges Related to Conceptual Issues

Despite decades of research in measuring segregation, we are still in the midst of defining segregation. While deriving a generally agreeable definition of segregation in unlikely to happen in sight, the needs to measure segregation continues. An immediate challenge is to identify some general principles in guiding the evaluation of segregation levels to avoid major pitfalls in the absence of a concrete definition of segregation. Based upon the review of literature provided above, several criteria seem to emerge.

Segregation involves more than one group. If the evaluating involves only one group, such evaluation would not capture segregation, regardless how segregation is defined operationally. Therefore, if one computes a measure reflecting the concentration of a group aspatially or spatially (e.g., percent Asian, or density of Hispanic), such concentration measure cannot reflect the relationship between groups and therefore are not effective to depict segregation levels. Similarly, if a spatial autocorrelation measure is used to evaluate the distribution of one population group, the result may show the extent that the group is spatially clustered, but does not necessarily reflect segregation. The use of binary group classifications such as group x vs. non-group x, or the proportion of population belonging to a group should be avoided if at all possible. Given the multi-racial-ethnic situations in many societies, the "other" group is so heterogeneous that its group identity may be problematic, and thus compromises the interpretation of the results.

Regardless if an agreeable conceptual definition of segregation is developed, the consensus is that an effective measure of segregation should be sensitive to the spatial arrangement of population distributions. In other words, a segregation measure should be spatially dependent, i.e., when the spatial distribution of population changes, the segregation measure should be able to reflect the change. For instance, when populations in two different groups exchange their locations, an effective segregation measure should be able to capture this change.

Closely related to the above criterion is about the spatial nature of segregation, which implies some forms of spatial separation. Therefore, segregation measures should include some metrics to indicate the spatial separation between population groups. While using distance is an effective way to capture spatial separation, distance can be measured and expressed through multiple metrics beyond the simple Euclidean distance (a generalized formulation of distance is the Mahalanobis distance). On the other hand, reduce forms of distance based upon topological relations, including ordered neighbors, are possible options to reflect spatial separation. The spatial statistical literature is rich in exploring different representations of spatial relation, and it should serve as the reference in formulating spatial segregation measures. On the other hand, our discussion on segregation has been implicitly limited to the separation in geographical space, but segregation can be extended to study separation in other spaces, including social network space and virtual spaces. Given the fact that social media is an important facet of life in today's societies, formal evaluations of segregation beyond the geographical space have not yet been explored, but are deemed appropriate and interesting.

Formulation of a measure often has to take into account the available data. Ecological or aggregated data, such as those provided by censuses have been the main source of data supporting segregation measurement. However, such ecological data assign individuals to areal units, separating populations into discrete spaces, while in reality the discrete spatial separation may not exist. Therefore, spatial segregation measures should not be affected by such artificial spatial discretization of space, but accounting for the interaction of people across unit boundaries when ecological data are used. This criterion may not be applicable to non-ecological data when location information of individual observations is available.

While spatial measures are appropriate for measuring segregation, they also carry a price. Not only that they are usually more complicated to compute, their executions often need GIS. Like any true spatial measures, spatial segregation measures have another methodological concern – boundary effect. The boundary has to be defined for a study region. Aspatial studies would typically exclude areas outside the boundary of the study region, not considering their presence. Such practices assume that units along the study region boundary have no relationships to units outside of the region, or the spatial relations along the edge of the study region are truncated. As units on the other side of the region boundary are removed from the study, units along the boundary will receive biased measurements due to ignoring the relations with units outside the boundary. The boundary effect has been discussed in the spatial statistics literature (e.g., Griffith and Amrhein 1983; Griffith 1985), but no solution is generally accepted. A simple, but not always feasible one is to include units beyond the study region to create a buffer region to reduce the boundary effect. However, how large the buffer region would be sufficient to reduce the boundary effect to an "acceptable" level is often difficult to judge. Also, if the study region boundary is a real rigid boundary, such as a coastline, water body, or an international border that prohibit movement or interaction of populations to the other side, creating a buffer region will not be feasible.

So where are we in the business of measuring segregation? Despite decades of effort in developing both aspatial and spatial measures of segregation, the empirical exercises by Massey and Denton, and later Johnston and his team, and the exercises such as those by Reardon and Firebaugh (2002), Reardon and O'Sullivan (2004) and Gorard and Taylor (2002) that studied the properties of measures, we have not been able to put a finger on a generally agreeable definition of segregation. Such discussions in the past two decades have been revolving around the five dimensions of segregation based upon empirical evidences. As point out in this review earlier, the meaning of some of the dimensions are problematic, imprecise and even confusing. Now may be the time to think out of the boxes provided by Massey and Denton, but focus on the real meaning of segregation, seeking strong conceptual bases, particularly along the spatial dimension of segregation.

5.6.2 Challenges Related to Data

As argued briefly above that measurement methodology and data are interdependent. Every measure requires particular types of data or data in certain formats. Some measures may be more flexible to accommodate multiple data types, but some may be more restrictive. Therefore, in order use a particular measure, appropriate data have to be available. Over the past several decades in the development of segregation measures, most measures rely on aggregated or ecological data, partly because censuses have been providing such data relatively consistently and reliably over time. An earlier section in this review has alluded to the development trend of increasing spatial resolution of observations for which values of segregation are computed. In addition to the general interest of comparing segregation levels across cities or metropolitan regions, we now are interested in comparing segregation levels at the neighborhood scale and the segregation levels experienced by individuals. Such extensions of inquiry along the spatial spectrum of observation units are also partly supported by the availability of new types of data. Neighborhood level data are still mostly ecological at reasonably high spatial resolutions. For instance, U.S. provides census data about the demographic characteristics at the street block level. In the U.K., the smallest census enumeration unit is the output area with an average of 400 people for England and Wales, and the 100-m grid cell for Northern Ireland. At the individual level, data with location information are now becoming more available, but such data are likely survey-type data rather than census-type data. In addition, proprietary data held by the private sector offer another valuable source, but are more difficult to obtain.

Several changes in the "data landscape" may post challenges, but they also provide opportunities to the use and development of segregation measurement. Along the aggregated-ecological data direction, a major change in the U.S. Census has been the abandoning of the long-form after the 2000 Census. The long form was used to sample about 16–17 % of U.S. population in previous censuses, gathering detailed socio-demographic and housing data. Data provided by the long-form have been the workhorses of many socioeconomic analyses in the U.S. for decades. As the long-form was abandoned after the 2000 Census, the American Community Survey (ACS) took its place to collect similar types of data continuously, but from a much smaller sample (visit http://www.census.gov/acs/www/ for more information about sample sizes and data products). Currency is the strength of ACS data with the trade-off of lower accuracy. Due to smaller sample sizes, the errors in ACS data are much larger than those from the long form data (such as Summary Files 3 and 4 in 2000 and earlier censuses). When earlier census data were used in segregation analyses, regardless if the data were from the long-form or short-form, data were assumed to be quite reliable, although they were not without errors. Thus, measuring segregation, like many socioeconomic analyses using census data, never takes into consideration the errors in data. However, this "error-free" assumption is no longer justifiable when using ACS data in general and for segregation analysis, as error for some ACS estimates could be substantial (see Sun and Wong 2010; Wong and Sun 2013 for some discussions). Therefore, with the availability of more current population data to support segregation studies, a new challenge is to include error information in evaluating segregation levels.

Despite the data quality issue, the ACS is likely the largest scale continuous survey measurement program and offers an improvement on the temporal granularity of gathering socioeconomic data of the population. Along the spatial granularity dimension, individual-level data are used more often now in socioeconomic research although the operational scales of these measurement programs are much smaller than ACS, and these data are often at the local or regional level. As there are numerous surveys of this type with individual information, a comprehensive discussion of these surveys will be quite impossible. Therefore, only a few directions are covered in the following discussions.

To a large extent, censuses are surveys in nature. In the U.S., the large scale population survey conducted by the Church of Jesus Christ of Latter-Day Saints in 1880 was similar to a census (Goeken et al. 2003). This database included individual records with some pertinent demographic information. Based upon these historical individual records, the Urban Transition Historical GIS Project (Logan et al. 2011) put more geography into the dataset by geocoding about five millions individuals in 39 cities, according to their addresses. Analyzing these individual-level records by a spatial association measure for categorical data, Páez et al. (2012, 2014) evaluated the segregation of different population groups at the micro or local level in four cities: Newark in New Jersey, Albany and Buffalo in New York, and Cincinnati in Ohio. This individual-level dataset is unique, but nonetheless, provides a possible data source to investigate segregation in the past.

On other hand, transportation studies and planning has a general need to understand the travel behavior of the population, including where people reside, where they visit, by what means do they travel, and for how long. To collect such information, travel diary surveys are often used, asking a sample of subjects to record their travel activities in a specific day. Such travel diary data have been used extensively to understand population mobility, transportation needs, and activity patterns. Wong and Shaw (2011) first demonstrated that such data gathered for the southern Florida region can be used to assess segregation at the individual level, based upon the activity space construct and concepts in contact theory. More recently, Farber et al. (2012) utilized a similar type of data for the Greater Montreal region to assess the interaction between linguistic groups to measure segregation. These studies demonstrate great potential of using such individual-level data for segregation analysis. However, these data definitely are different from census-based data as they are surveys including only small segments of the entire population. Their reliability and representativeness are issues of concern. In addition, given the individual nature of the data, existing measures designed for aggregated-ecological data may not be applicable. Either existing measures have to be modified or new measures have to be developed to utilize individual-level data.

Besides the increasing use of surveys to gather individual-level data, a relatively new source of individual-level data fueling a new wave of social research are generated from social media. In the social media environments, users are in essence providing data, either data about themselves (e.g., describing where they have been), or what they have sensed (e.g., witnessing certain events or incidents). From the spatial science perspective, these users often provide geographic information, either explicitly or implicitly. Using these data for geographical studies has been a focus of research under the umbrella topic of volunteered geographic information (VGI) (e.g., Elwood et al. 2012). Using social media data for geographic research has many challenges, both technically and conceptually, but undeniably, such data are rich in (geographical) content and are relatively current (e.g., Xu et al. 2013). To what extent social media can be used for measuring segregation has not been explored yet, but should be an interesting direction to pursue.

References

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.

Bavaud, F. (1998). Models for spatial weights: A systemic look. *Geographical Analysis, 30*, 153–177.

Bell, W. (1954). A probability model for the measurement of ecological segregation. *Social Forces, 32*, 357–364.

Brown, L. A., & Chung, S.-Y. (2006). Spatial segregation, segregation indices and the geographical perspective. *Population, Space and Place, 12*(2), 125–143.

Cortese, C., Falk, F., & Cohen, J. (1976). Further considerations on the methodological analysis of segregation indices. *American Sociological Review, 41*(4), 630–637.

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indices. *American Sociological Review, 20*, 210–217.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers, 102*(3), 571–590.

Farber, S., Páez, A., & Morency, C. (2012). Activity spaces and the measurement of clustering and exposure: A case study of linguistic groups in Montreal. *Environment and Planning A, 44*, 315–332.

Feitosa, F. F., Camara, G., Monteiro, A. M. V., Koschitzki, T., & Silva, M. P. S. (2007). Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science, 21*, 299–323.

Fotheringham, A. F. (1997). Trends in quantitative methods I: Stressing the local. *Progress in Human Geography, 21*(1), 88–96.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*, 189–206.

Goeken, R., Nguyen, C., Ruggles, S., & Sargent, W. (2003). The 1880 United States population database. *Historical Methods, 32*, 27–34.

Gorard, S., & Taylor, C. (2002). What is segregation? A comparison of measures in terms of 'strong' and 'weak' compositional invariance. *Sociology: The Journal of the British Sociological Association, 36*, 875–895.

Grady, S. C. (2006). Racial disparities in low birthweight and the contribution of residential segregation: A multilevel analysis. *Social Science and Medicine, 63*(12), 3013–3029.

Griffith, D. A. (1985). An evaluation of correction techniques for boundary effects in spatial statistical analysis: Contemporary methods. *Geographical Analysis, 17*(1), 81–88.

Griffith, D. A. (1996). Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In S. L. Arlinghaus & D. A. Griffith (Eds.), *Practical handbook of spatial statistics* (pp. 65–82). Boca Raton: CRC Press.

Griffith, D. A., & Amrhein, C. G. (1983). An evaluation of correction techniques for boundary effects in spatial statistical analysis: Traditional methods. *Geographical Analysis, 15*(4), 352–360.

Jahn, J., Schmid, C. F., & Schrag, C. (1947). The measurement of ecological segregation. *American Sociological Review, 12*, 293–303.

Jakubs, J. F. (1981). A distance-based segregation index. *Journal of Socio-Economic Planning Sciences, 15*, 129–136.

Johnston, R., Poulsen, M., & Forrest, J. (2007). Racial and ethnic segregation in U.S. Metropolitan areas, 1980–2000: The dimensions of segregation revisited. *Urban Affairs Review, 42*, 479–504.

Johnston, R., Poulsen, M., & Forrest, J. (2009). Measuring ethnic residential segregation: Putting some more geography in. *Urban Geography, 30*, 91–109.

Krupka, D. J. (2007). Are big cities more segregated? Neighborhood scale and the measurement of segregation. *Urban Studies, 44*(1), 187–197.

Lee, C.-M., & Culhane, D. P. (1998). A perimeter-based clustering index for measuring spatial segregation: A cognitive GIS approach. *Environment and Planning B, 25*, 327–343.

Lieberson, S. (1981). An asymmetrical approach to segregation. In C. Peach, V. Robinson, & S. Smith (Eds.), *Ethnic segregation in cities* (pp. 61–82). London: Croom-Helm.

Logan, J. R., Jindrich, J., Shin, H., & Zhang, W. (2011). Mapping America in 1880: The urban transition historical GIS project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 44*, 49–60.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*, 281–315.

Massey, D. S., White, M. J., & Phua, V. C. (1996). The dimensions of segregation revisited. *Sociological Methods & Research, 25*, 172–206.

Morgan, B. S. (1975). The segregation of socioeconomic groups in urban areas. *Urban Studies, 12*, 47–60.

Morgan, B. S. (1983a). An alternate approach to the development of a distance-based measure of racial segregation. *American Journal of Sociology, 88*, 1237–1249.

Morgan, B. S. (1983b). A distance-decay interaction index to measure residential segregation. *Area, 15*, 211–216.

Morrill, R. L. (1991). On the measure of geographic segregation. *Geography Research Forum, 11*, 25–36.

Newby, R. G. (1982). Segregation, desegregation, and racial balance: Status implications of these concepts. *The Urban Review, 14*, 17–24.

O'Sullivan, D., & Wong, D. W. (2007). A surface-based approach to measuring spatial segregation. *Geographical Analysis, 39*, 147–168.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127–144). London: Pion.

Östh, J., Malmberg, B., & Andersson, E. (2014). Analysing segregation with individualized neighbourhoods. In C. D. Lloyd, I. Shuttleworth, & D. W. Wong (Eds.), *Social-spatial segregation*. Bristol: Policy Press.

Páez, A., Ruiz, M., López, F., & Logan, J. (2012). Measuring ethnic clustering and exposure with the q statistic: An exploratory analysis of Irish, Germans, and Yankees in 1880 Newark. *Annals of the Association of American Geographers, 102*, 84–102.

Páez, A., Ruiz, M., López, F., & Logan, J. (2014). The micro-geography of segregation: Evidence from historical US census data. In C. D. Lloyd, I. Shuttleworth, & D. W. Wong (Eds.), *Social-spatial segregation*. Bristol: Policy Press.

Peach, C. (1996). The meaning of segregation. *Planning Practice and Research, 11*(2), 137–150.

Poulsen, M. F., Johnston, R. J., & Forrest, J. (2002). Plural cities and ethnic enclaves: Introducing a measurement procedure for comparative study. *International Journal of Urban and Regional Research, 26*(2), 229–243.

Poulsen, M. F., Johnston, R. J., & Forrest, J. (2010). The intensity of ethnic residential clustering: Exploring scale effects using local indicators of spatial association. *Environment and Planning A, 42*, 874–894.

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology, 32*, 33–67.

Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology, 34*, 121–162.

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography, 45*(3), 489–514.

Reardon, S. F., Farrell, C. R., Matthews, S. A., O'Sullivan, D., Bischoff, K., & Firebaugh, G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation: 1990–2000. *Social Science Research, 38*, 55–70.

Reiner, T. A. (1972). Racial segregation: A comment. *Journal of Regional Science, 12*(1), 137–148.

Sakoda, J. N. (1981). A generalized index of dissimilarity. *Demography, 18*, 245–250.

Shuttleworth, I. G., Lloyd, C. D., & Martin, D. J. (2010). Exploring the implications of changing census output geographies for the measurement of residential segregation: The example of Northern Ireland 1991–2001. *Journal of the Royal Statistical Society, Series A, 174*(1), 1–16.

Sun, M., & Wong, D. W. S. (2010). Incorporating data quality information in mapping the American Community Survey data. *Cartography and Geographic Information Science, 37*(4), 285–300.

Taeuber, K. E., & Taeuber, A. F. (1965). *Negroes in cities: Residential segregation and neighborhood change*. Chicago: Aldine.

Taeuber, K. E., & Taeuber, A. F. (1976). A practitioner's perspective on the index of dissimilarity. *American Sociological Review, 41*, 884–889.

Taylor, C., Gorard, S., & Fitz, J. (2010). The modifiable areal unit problem: Segregation between schools and levels of analysis. *International Journal of Social Research Methodology, 6*(1), 41–60.

Wang, Y. (2012). Decomposing the entropy index of racial diversity: In search of two types of variance. *Annal of Regional Science, 48*, 897–915.

White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology, 88*, 1008–1018.

White, M. J., Kim, A. H., & Glick, J. E. (2005). Mapping social distance ethnic residential segregation in a multiethnic Metro. *Sociological Methods Research November, 34*(2), 173–203.

White, K., Borrell, L. N., Wong, D. W., Galea, S., Ogedegbe, G., & Glymour, M. M. (2011). Racial/ethnic residential segregation and self-reported hypertension among U.S.- and foreign-born blacks. *American Journal of Hypertension, 24*(8), 904–910.

Winship, C. (1978). The desirability of using the index of dissimilarity or any adjustment of it for measuring segregation. *Social Forces, 57*, 717–721.

Wong, D. W. S. (1993). Spatial indices of segregation. *Urban Studies, 30*, 559–572.

Wong, D. W. S. (1996). Enhancing segregation studies using GIS. *Computers, Environment and Urban Systems, 20*(2), 99–109.

Wong, D. W. S. (1997). Spatial dependency of segregation indices. *The Canadian Geographer, 41*(2), 128–136.

Wong, D. W. S. (1998). Measuring multi-ethnic spatial segregation. *Urban Geography, 19*(1), 77–87.

Wong, D. W. S. (2002). Modeling local segregation: A spatial interaction approach. *Geographical and Environmental Modelling, 6*(1), 81–97.

Wong, D. W. S. (2003). Spatial decomposition of segregation indices: A framework toward measuring segregation at multiple levels. *Geographical Analysis, 35*(3), 179–194.
Wong, D. W. S. (2004). Comparing traditional and spatial segregation measures: A spatial scale perspective. *Urban Geography, 25*(1), 66–82.
Wong, D. W. (2005). Formulating a general spatial segregation measure. *The Professional Geographer, 57*(2), 285–294.
Wong, D. (2009). The Modifiable Areal Unit Problem (MAUP). In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE handbook of spatial analysis* (pp. 105–123). London: Sage.
Wong, D. W. S. (2011). Exploring spatial patterns using an expanded spatial autocorrelation framework. *Geographical Analysis, 43*(3), 327–338.
Wong, D. W. S. (2014). Using a general spatial pattern statistic to evaluate spatial segregation. In C. D. Lloyd, I. Shuttleworth, & D. W. Wong (Eds.), *Social-spatial segregation*. Bristol: Policy Press.
Wong, D. W. S., & Shaw, S.-L. (2011). Measuring segregation: An active-space approach. *Journal of Geographical Systems, 13*(2), 127–145.
Wong, D. W., & Sun, M. (2013). Handling data quality information of survey data in GIS: A case of using the American community survey data. *Spatial Demography, 1*(1), 3–16.
Wong, D. W. S., Lasus, H., & Falk, R. F. (1999). Exploring the variability of segregation index D with scale and zonal systems: An analysis of thirty U.S. cities. *Environment and Planning A, 31*, 507–522.
Xu, C., Wong, D. W., & Yang, C. (2013). Evaluating the "Geographical Awareness" of individuals: An exploratory analysis of Twitter data. *Cartography and Geographic Information Science, 40*(2), 103–115.

Chapter 6
Demography Is an Inherently Spatial Science

John R. Weeks

6.1 Introduction

Demography is, by its very nature, concerned with people in places, although the history of the discipline over time reveals a struggle between the desire to find universal principles (such as the original model of the demographic transition) and the recognition that spatial variation is itself a universal principle. Demography is in the process of evolving from a spatially aware science to a spatially analytic science, and this book is part of that evolution. In this chapter I first offer a general framework for the application of spatial analysis to demographic research as a way of integrating and better understanding the different transitional components of the overall demographic transition. I then illustrate tools of spatial demography by applying them to an analysis of demographic change in the West African country of Ghana, with an added focus on Accra, the country's capital city.

6.2 The Demographic Transition Is Really a Suite of Transitions

Although it has dominated demographic thinking for the past half century, the demographic transition theory actually began simply as a description of the demographic changes that had taken place in the advanced nations over time. In particular, it described the temporal shift from high birth and death rates to low birth and death rates, with an interstitial spurt in growth rates leading to a larger population at

J.R. Weeks (✉)
International Population Center, Department of Geography,
San Diego State University, San Diego, CA, USA
e-mail: john.weeks@sdsu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_6

the end of the transition than there had been at the start. As it became clear that this was a global phenomenon, unique in human history, modernization theory was invoked to offer an explanation. Modernization theory is based on the idea that in premodern times human society was generally governed by "tradition," and that the massive economic changes wrought by industrialization forced societies to alter traditional institutions: "In traditional societies fertility and mortality are high. In modern societies fertility and mortality are low. In between, there is demographic transition" (Demeny 1968:502). Death rates decline as the standard of living improves, and birth rates almost always decline a few decades later, eventually dropping to low levels. It was argued that the decline in the birth rate typically lags behind the decline in the death rate because it takes time for a population to adjust to the fact that mortality really is lower, and because the social and economic institutions that favor high fertility require time to adjust to new norms of lower fertility that are more consistent with the lower levels of mortality. Since most people value the prolongation of life, and because key drivers of lower mortality are communal in nature (clean water, sanitation, immunization programs, etc.) it is not too hard to lower mortality, but the reduction of fertility is contrary to the established norms of societies that have required high birth rates to keep pace with high death rates. Such norms, which are implemented at the individual level (e.g., couples having unprotected sex) are not easily changed, even in the face of poverty.

Birth rates eventually decline, it was argued, as the importance of family life is diminished by industrial and urban life, thus weakening the pressure for large families. Large families are presumed to have been desired because they provided parents with a built-in labor pool, and because children provided old-age security for parents. The same economic development that lowered mortality is theorized to transform a society into an urban industrial state in which compulsory education lowers the value of children by removing them from the labor force, and people come to realize that lower infant mortality means that fewer children need to be born to achieve a certain number of surviving children (Easterlin 1978). Finally, as a consequence of the many alterations in social institutions, "the pressure from high fertility weakens and the idea of conscious control of fertility gradually gains strength" (Teitelbaum 1975:421).

Modernization thus focuses on the economic drivers of demographic change. It turns out that cultural factors play a role, as well. This idea first emerged from analyses being done as part of Princeton University's European Fertility Project, in which it was discovered that the decline of fertility in Europe occurred in the context of widely differing social, economic, and demographic conditions. It thus became apparent that economic development was a sufficient cause of fertility decline, but not a necessary one (Coale 1973). For example, many provinces in Europe experienced a rapid drop in their birth rate even though they were not very urban, infant mortality rates were high, and a low percentage of the population was in industrial occupations. The data suggest that one of the more common similarities in those areas that had undergone early fertility declines was the rapid spread of "secularization," which is an attitude of autonomy from otherworldly powers and a sense of responsibility for one's own well-being (Lesthaeghe 1977; Leasure 1982; Norris and Inglehart 2004). It is associated with an enlightened view of the world—a break from traditional ways of thinking and behaving.

In retrospect, we can see that the innovation of fertility declines in Europe provided a nearly classic example of spatial autocorrelation, of Tobler's First Law of Geography that everything is related to everything else, but near things are more related than distant things (Tobler 1970, 2004). Were it not for spatial autocorrelation, fertility might have declined in isolated settings, but the decline would not have spread as it did. It turns out that all three demographic processes—mortality, fertility, and migration—exhibit spatial autocorrelation, illustrated by almost any map of the countries of the world (or regions within countries) showing differences in these demographic phenomena. Nearly all of these differences are associated in some way or another with human culture (Davis 1949). Even when it seems as though the environment may be important, as in differences in health levels according to whether a person lives in a slum or not, at root most differences are cultural in terms of how some neighborhoods are organized compared to others. Even when it seems as though biology may be important, as in differences in hypertension or obesity, at root most differences are cultural in terms of diet, exercise, and access to modern health care.

Since culture underlies most aspects of demography, if we can understand why some places have different cultures than others, we are in a good position to understand spatially varying levels of mortality, fertility, and migration. Culture is a very complex concept, so we typically study it in terms of the social institutions (e.g., family, religion, and economy) created by humans and forming the structure of society. Humans are social by nature. Indeed, humans are rarely able to survive on their own. Within each social group, culture arises initially in response to solutions that people derive for the problems and issues of everyday life: how to communicate with one another, where to live, what to eat and how to eat it (institution of the economy), how to ensure that children are reproduced, reared and taught how to behave (institution of the family), how to encourage good behavior and punish bad behavior (institution of government), how to explain things that do not have obvious answers (institution of religion). Over the long course of human history, different groups of people have derived different solutions, although many such solutions are similar to those of other groups. Furthermore, contact between groups can lead to the diffusion of innovations, either through coercion or by choice. There are clearly spatial factors at work.

Historians tend to agree that, of all the innovations and inventions created by humans, the single most important one was the movable-type printing press of Gutenberg (Ferguson 2011). This led to books being accessible in a way that was almost unimaginable, which encouraged writing, which encouraged critical thinking, which led to science, which led to the control of death—an innovation whose spatial diffusion has dramatically altered the world, one family, one community, one society at a time. From this perspective, improvements in mortality were not so much dependent upon modernization as they were integral to it. The same scientific ways of viewing the world that led to the industrial revolution also led to the science of death control.

This more cultural view of demographic history, emphasizing enlightened thinking and science instead of just economic changes, began most clearly with Kingsley Davis's theory of demographic change and response (Davis 1963). He started out with the basic question of how, and under what conditions, can a mortality decline lead to a fertility decline? The answer begins with the realization that a mortality decline is first experienced by families. Death control historically has benefited children first and foremost, so when mortality declines it means that more children survive through adulthood, putting greater pressure on family resources, and people have to reorganize their lives in an attempt to relieve that pressure; that is, people respond to the demographic change. Davis argued that the response that individuals make to the population pressure created by more members joining their ranks is determined by the means available to them. A first response, non-demographic in nature, is to try to increase resources by working harder—longer hours perhaps, a second job, and so on. In this way, a decline in mortality can actually lead to greater economic productivity, rather than the other way around, just as Danish economist Ester Boserup argued (Boserup 1981). If that is not sufficient or there are no such opportunities for economic growth (as there often are not in subsistence agricultural communities), then migration of some family members (typically, at least until recently, unmarried sons or daughters) is the easiest demographic response. Migration is, of course, the option that people have been using forever, helping to explain the spatial spread of humans across the globe.

The theory of demographic change and response alerts us to the fact that the three basic demographic processes of mortality, fertility, and migration are intimately tied up together. Indeed, I have argued for years that the demographic transition, which provides the organizing theoretical framework for most demographic research, is really a complex set of transitions, each of which draws upon expertise and perspectives in differing social science and health-related disciplines (Weeks 2012). While these transitions arise from our observations about the world, that is how science unfolds, and the transitions provide us with the kinds of testable hypotheses (expectations) about social inequalities and changes that we associate with middle range theory. The suite of transitions that comprise the overall demographic transition is illustrated in Fig. 6.1. The process almost always begins with the health and mortality transition, which is the shift over time from high death rates with deaths clustered at the younger ages and caused largely by communicable diseases, to low death rates with deaths clustered at the older ages and caused largely by degenerative diseases. This is probably the most transformative change ever to happen to human society, and comes directly out of the scientific advances put in motion over the last two hundred or so years by the Enlightenment, as noted above. The health and mortality transition initiates a chain of other demographic events, all of which impact each other and are, at the same time, influenced by external societal events. The way in which these transitions take place then shapes what societies can and will be.

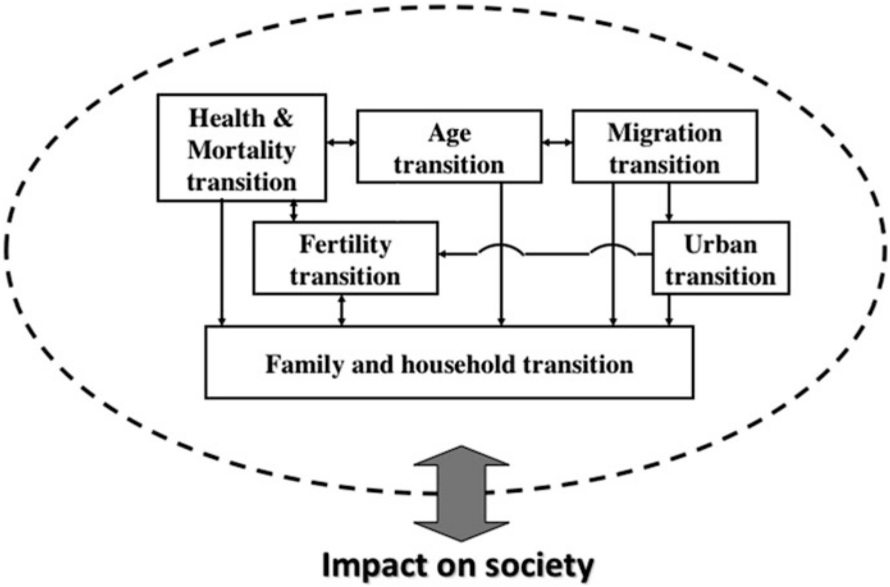

Fig. 6.1 The transitions that comprise the way in which demography and society impact each other (Source: Adapted with permission from Weeks 2012: Figure 3.4)

The fertility transition represents the change from high fertility levels over which people have relatively little direct control to low fertility over which people have considerable control. Control is also tied closely to science, since methods used before the Enlightenment—especially withdrawal (which is mentioned in the Old Testament), herbal remedies to prevent conception, and attempts at abstinence, have very low rates of efficacy. For most of human history, high death rates were either matched by high birth rates, or else a society disappeared over time because at the historical level of 20 years life expectancy the average woman must give birth to six babies in order to ensure that two will survive to adulthood. With declining death rates, historically unprecedented numbers of children survive to adulthood and society has to respond in multiple ways, including (at least eventually) with an increase in the effectiveness of fertility control. When life expectancy gets past the mid-50s, it takes scarcely two children being born on average to each woman to ensure that two children will survive to adulthood. Of course, the longer couples delay in lowering fertility as mortality declines, the more rapidly will the population grow, and the greater and more varied will the societal adaptations have to be.

The migration transition is initially the predictable response to population growth in rural areas, where there are not enough jobs to go around as the birth rate begins to exceed the death rate and more children survive to adulthood and need a job. The theory of demographic change and response suggests that people try increasing productivity, but when those possibilities are exhausted, people are motivated by necessity to seek opportunity elsewhere, increasingly over time in urban places, thus unleashing the urban transition, in which a population moves from being largely rural to being largely urban.

The urban transition is especially important because it is in this environment that cultural changes are most apt to take place, as Weeks et al. (2004:75) have noted:

> If we accept this idea of culture, then we can see that culture is bound to be highly spatial in its nature because it is easier to copy than to invent (the essence of diffusion) and people are likely to copy solutions to their problems from neighbors: the fewer and the less diverse your neighbors, the fewer options you have from which to choose. The city is the fount of innovation, including that with respect to human reproduction, precisely because it brings together a greater diversity of people and their different solutions to life's problems than will typically exist in a small rural village. In the latter places, it is much more likely that a group's solutions will become reified—perhaps justified as having been derived from a super-natural power or thought of as having been inherited genetically. This promotes resistance to change, including change in family structure, gender relations, and reproductive behavior.

The age transition is a predictable result of changes in mortality, fertility and migration, in which high mortality and high fertility produce a very young age structure that is pyramid-shaped, with a broad base of young people narrowing to a very small number of older people. Initially the declines in mortality actually broaden that young base because a greater fraction of children surviving acts like an increase in the birth rate. However, as fertility declines in the face of lower mortality, it initially generates a dent in the youngest ages as fewer children are born, and this then generates bulges in the young adult ages, leading eventually to a barrel-shaped age structure (sometimes called the "demographic dividend"). If the birth and death rates wind up equaling each other but at very low rates, the resulting age structure has nearly the same number of people at each age until the very oldest ages, when people die off at an advanced age. Migration adds its own special twist to the age structure because migrants are disproportionately young adults. When they leave a place of origin, their departure creates a dent in the young adult ages, while creating a youth bulge in the place of origin. There are long-term consequences, as well, since young adulthood is also the time of reproduction, so places of out-migration will be on a trajectory to have fewer births than would otherwise be expected, whereas the places of destination will get a birth boost from the migrants. The age transition is so important that I call it the "master transition," because the changes in the number and percent of people at each age force societies to adapt in some way or another.

The family and household transition is an often-neglected part of the demographic transition, yet it is the aspect of the demographic transition that people are often most concerned about, without realizing the sources of what amount to dramatic shifts in how society looks and works. This transition represents the change from complex forms of family and household structure when mortality and fertility are both high (because families are being constantly broken apart by death and then new families created on the backs of the old ones to keep the community going), to less variability in the middle of the transition (when child and maternal mortality have dropped, but before later adult mortality has declined substantially, and when most people are still in rural environments) to new forms of complexity when both fertility and mortality are low, and most people are living in urban areas, living long lives, and thus probably not still co-resident with their small number of children. Finally, of course, there is the overall transition in population size that occurs when mortality declines sooner than fertility (the usual pattern in the demographic transition) from which massive changes follow with respect to resource utilization and allocation, not just at the local level, but at the regional and global level.

6.3 The Spatial Nature of the Transitions

How a given society responds to each of these transitions is related to the spatial context in which that society exists. Each of these interrelated aspects of demographic change have spatial (and temporal) components which, when understood, add to our knowledge of how and why these transitions occur and what their impact will be. There are three spatial elements, in particular, that play a role in the different timing and pattern of each of the transitions shown in Fig. 6.1. These are (1) space—demographic changes vary across a region as a function of differences in characteristics such as cultural, economic, and political history, natural environment, and built environment (infrastructure); (2) place ("neighborhood context"—broadly defined—matters when it comes to virtually all aspects of human behavior); and (3) scale (some things are more local in their effects than are others) (see John Logan's Chap. 2 in this volume for more on the differences between space, place, and scale.)

The fact that demography is spatial by nature means that much, if not most, of the demographic research that is conducted has a spatial "awareness," even if relatively little—albeit a growing segment—of it yet engages spatial "analysis" in any formal sense. Spatially aware research grasps the essence of Tobler's famous First Law of Geography in that demographic behavior can, for example, be expected to differ by geographic region; population characteristics and change are different in urban than in rural places; countries in sub-Saharan Africa with a high proportion of Muslims have lower HIV/AIDS prevalence rates than predominantly non-Muslim nations; and East Asian countries have experienced a different fertility transition than South Asian countries.

Migration research has historically been the staple of population geographers, dating at least back to Ravenstein's classic analysis of "The Birthplaces of the People and the Laws of Migration," which was built from data in the 1871 census data for Britain (Ravenstein 1876). Migration has a built-in spatial awareness because the analyses focus on the places from which migrants come and the places to which they go, and the networks that are created by the movements of people. Migration matrices and multi-regional life tables have been created as tools that increase our quantitative understanding of these changes involved in migration.

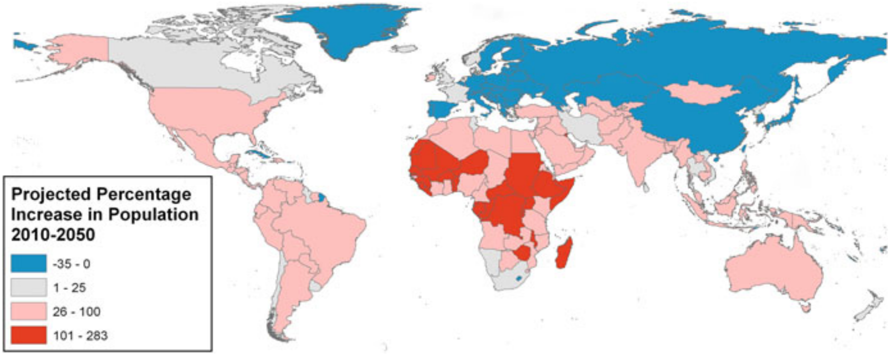

Fig. 6.2 Projected percentage increase in total population size between 2010 and 2050 (Source: United Nations Population Division (2011): http://esa.un.org/unpd/wpp/unpp/panel_population.htm (accessed 2013))

Figure 6.2 provides the kind of demographic visualization that instantly reminds us of the spatial component of demography. This map shows the countries of the world according to the projected rate of population growth between 2010 and 2050, based on projections prepared by the United Nations Population Division (2011). Throughout Europe populations are projected to decline or at least grow only very slowly, while at the other extreme the highest rates of growth are expected to be in sub-Saharan Africa. This is the area of the world that is growing in population quickly, yet almost invisibly. The latest UN projections suggest that Nigeria could be the third most populous nation by 2050, surpassing the United States, and Fig. 6.2 shows that most of Nigeria's neighbors are growing at a faster rate than Nigeria, although they are not building on quite such a large population base as Nigeria's. This pattern of population change almost certainly will alter the world's geopolitical structure. But visual awareness of space and place is not quite the same as spatial demographic analysis because it is not typically associated with underlying theories and hypotheses about spatial patterns that are designed to be tested for their specific spatial content. This is largely due to the recency with which the tools of GIScience and spatial statistics have become available.

6.4 A General Framework for Spatial Demographic Analysis

Spatial analysis can be defined as a quantitative data analysis in which the focus is on the role of space and which relies on explicitly spatial variables in the explanation or prediction of the phenomenon under investigation (Cressie 1993; Chou 1997). Spatial analysis in the social sciences tests theories that where you are makes a difference in social attitudes and behavior, and that observed differences in the social world are not distributed in a spatially random pattern. Cressie (1993) argues that the classical, nonspatial data analysis should actually be seen as a special case of spatial data analysis. Viewed in this way, the underlying logic is that each random variable (z) is associated with locational attributes (x and y). In spatial data analysis, the researcher uses spatial statistics to glean information from the x and y coordinates, whereas in classical statistical analysis the researcher ignores those coordinates (often not even realizing that they might exist). More to the point, in classical statistical analysis, the locational attributes are considered to be a nuisance, rather than representing useful information. Spatial attributes are things to be gotten rid of, or controlled for, whereas in spatial data analysis they become objects of investigation.

There are two key and interconnected aspects of spatial patterns that we must account for: (1) spatial dependence; and (2) spatial heterogeneity. Spatial dependence takes us back to Tobler's First Law of Geography, that near things may be correlated with one another because they are spatially contiguous. Proximity is thus a predictor of some aspects of behavior. For example, everywhere we go in the world, fertility is lower among better educated women than among less well-educated women. But less well-educated women may have fewer children than you might otherwise expect if they live near better educated women because of the diffusion of attitudes about family size and knowledge of family planning (see, for example, Weeks et al. 2004).

Spatial heterogeneity (sometimes known as spatial non-stationarity) refers to the situation where associations among variables are different in some places than they are in other places. This is also known as a situation of "spatially varying coefficients" (Fotheringham et al. 2002). Spatial heterogeneity is a special case of spatial dependence, in which not only are near things more highly correlated than distant things, but the strength (e.g., strong or weak) and perhaps even the direction of the relationship (e.g., positive or negative) varies from place to place. Spatial dependence does not always include spatial heterogeneity, but spatial heterogeneity always involves spatial dependence. A good example of spatial heterogeneity can be found by going back to the relationship between fertility and education. While this negative relationship holds almost everywhere in the world, in sub-Saharan Africa a well-educated woman is apt to have more children than a similarly well-educated woman in Europe, just as a poorly educated woman in sub-Saharan Africa is apt to have more children than a similarly poorly educated woman in Europe. Knowing a woman's level of education will not let you automatically predict the number of children she has; rather it tells you that she likely has fewer children than a less well educated women in her part of the world. The explanation has to be sought in local cultural norms, which tend to be place-specific, meaning that they have a spatial component.

The comments about spatial dependence (also known as spatial autocorrelation) apply as well to temporal autocorrelation (things that are close to one another temporally are more likely to be similar than things that are more temporally distant). Econometricians have developed autoregressive models to account for the temporal autocorrelation that is typically found in time-series data that comprise the backbone of much of economic analysis. Time can be thought of a disturbance to be controlled for, as well as an effect to be studied. Spatial analytic methods are, for the most part, derived from these econometric models in which space is substituted for time (Anselin 1988).

In order to create a general framework for spatial analysis in demography, I draw upon the ideas of Star and Estes (1990) that spatial analysis can be divided into two "families": (1) analysis that is concerned with *local or neighborhood characteristics*; and (2) analysis that is concerned with *connections* between locations. In demographic research we can think of the local or neighborhood characteristics as representing aspects of the context (place) in which demographic decisions are made and demographic behavior is manifested. Spatial analysis then looks for place-specific factors that influence the behavior of otherwise similar people. At the same time, we are always cognizant of the fact that scale matters—how we define the size and/or boundaries of a place may influence our understanding of what's happening. The connections, on the other hand, relate to the kinds of networking and interaction that lead both to *diffusion* (the spread of ideas) and *dispersal* (the geographic redistribution of people). Spatial analysis then searches for the timing and direction of the connections, and seeks to understand their causes and consequences, both of which may be related to context at the local or neighborhood level.

Spatial demographic analysis, whether it be spatial context analysis or network linkages, incorporates key elements of the broader field of spatial analysis. Indeed, my definition of spatial demography is very simple: it represents the application of spatial concepts and statistics to demographic phenomena (see Weeks 2004; Voss 2007). Spatial demography is different from population geography because the latter has traditionally focused only on the mobility of people—the network part of analysis, whereas spatial demography looks at all spatial aspects of all demographic phenomena. It is different from applied demography for similar reasons—it is broader in scope than applied demography, which tends to focus on practical uses of demographic information, such as projections of school population, or the demographic characteristics of consumers of a particular type of product. Out of applied demography has emerged the subfield of geodemographics, which is a part of spatial demography, but which tends largely to focus on neighborhood characteristics—are people with similar tastes for consumer goods spatially clustered in the same neighborhoods? If so, we can achieve some efficiencies in marketing certain products to them.

Spatial demography thus looks at all aspects of demography—each of the suites in the overall demographic transition as shown in Fig. 6.1, and tries to sort out why we see the spatial patterns of the type shown in Fig. 6.2. The basic requirement for undertaking spatial demography is to have georeferenced data—demographic data with locational attributes attached to them. The spatial scale of georeferenced data may vary from region or country all the way down to a person's address or even their GPS tracks during a typical day. No matter the scale, the point is that each piece of information has some kind of geographic data associated with it. This may be a place at a given point in time (most common), or different places at different points in time, which may then translate into flows over time from one place to another. The kinds of data that might be georeferenced and thus be available for spatial demographic analysis include censuses, surveys, vital statistics, administrative data, and remotely sensed data. Let me illustrate some of these general concepts in more detail through an analysis of data for Ghana, a West African nation that, as can be seen in Fig. 6.2, now has a slower rate of population growth than most of its neighbors.

6.5 Spatial Demographic Concepts Applied to Ghana in West Africa

Ghana was the first sub-Saharan African country to gain full independence from Britain (in 1957) and despite episodes of military rule, has emerged as one of the more prosperous and stable democratic countries in a region that remains very poor and prone to civil war. Like all other Sub-Saharan countries, Ghana is experiencing rapid population growth and very rapid urbanization, and the future of the country depends very largely on economic, social, political and cultural development in its cities. For the past decade, a research team in which I have been involved has joined others striving to understand the changing population dynamics in Ghana, with a focus on its capital city, the burgeoning West African metropolis of Accra (Weeks et al. 2013b).

6.5.1 The Regional Context of Ghana

It is clear from Fig. 6.2 that Ghana is in the region of the world that has the highest overall rate of population growth, with much of that growth being funneled into cities, as people search for work in an ever-crowded, but still quite poor, region. Davis (2007:5–6) talks about the "vast West African conurbation rapidly coalescing along the Gulf of Guinea with Lagos (23 million by 2015 by one estimate) as its fulcrum. . .[and] a total of more than 60 million inhabitants along a strip of land 600 km, running east to west between Benin and Accra. Tragically, it probably will also be the biggest single footprint of urban poverty on earth."

Like its neighbors, Ghana on the eve of independence had high mortality (female life expectancy of 42 years and under-five mortality of 250 children per 1,000 live births) and high fertility (TFR = 6.4) with a very young age structure (45 % of the population under the age of 15) and a population that was largely rural (only 15 % urban), living in extended families, typically in compound-style housing. But, despite the very high mortality, the population in 1950 (which was only five million then) was growing at a rate of 2.6 % per year—a rate that would lead to a doubling of the population in only 27 years. Indeed, it was just about 27 years later, in the late

1970s, that Ghana's population was twice as large, at ten million. Under-five mortality is estimated to have been 350 per 1,000 in the mid-1930s (Weeks et al. 2013a), so the figure of 250 in 1950 was a substantial decline, and since fertility had not yet started to decline the population was growing.

As of 2010, Ghana as a country was moving slowly through the demographic transition. The population of 25 million is five times what it was in 1950, and the growth rate has dropped to 2 % (an implied doubling time of 35 years). Life expectancy is higher (61 years for females), under-five mortality is down to 78 per 1,000, fertility has dropped to 3.9 children, the age structure is not quite so young (39 % of the population under 15), and perhaps most dramatically from outward appearances, the country now has a majority of its population (52 %) living in urban places. Inwardly, though, the dramatic change has been the control of communicable disease that has allowed a much higher fraction of children to survive to adulthood. That's the good news. The bad news is that the drop in fertility has not matched the decline in mortality, so the population is growing rapidly, and only the urban places can offer much hope for jobs to young people trying to get a start in life. So, migration is towards the cities, which are largely in the south of the country, but even the northern city of Tamale has seen a huge expansion of population (Ghana Statistical Service 2012), lying as it does in the middle of the Northern Region, which has the highest rate of growth in the country.

6.5.2 The Regional Context Within Ghana

Ghana has progressed farther through the demographic transition than most of its near neighbors, so within West Africa there is demographic variability, as hinted at in Fig. 6.2. But we can better appreciate West African demography by looking at the variability within Ghana. An excellent source of georeferenced demographic data for a large number of developing countries is the set of Demographic and Health Surveys (www.measuredhs.com), administered by Measure DHS (based in the Washington, DC area) and funded largely by the US Agency for International Development, but conducted always in cooperation with the national statistical agency of each participating country. The sampling strategy is a multistage cluster probability sample, aimed at providing statistically significant results for the entire country, but also for defined administrative units within each country. For example, there are ten such areas, called regions, within Ghana. Figure 6.3 shows the spatial pattern, by region, of the total fertility rate (TFR) and the under-five mortality rate (U5MR) from the 2008 Ghana Demographic and Health Survey (GDHS), the most recent survey as of this writing. Digital boundary files for many of the countries in which Demographic and Health Surveys have been administered can be downloaded from the website of the United Nations' Secondary Administrative Level Boundary Project (SALB—http://www.unsalb.org/).

It is clear from Fig. 6.3 that there is considerable spatial variability in both TFR and U5MR in Ghana as of 2008. There is a general north-south pattern, in which

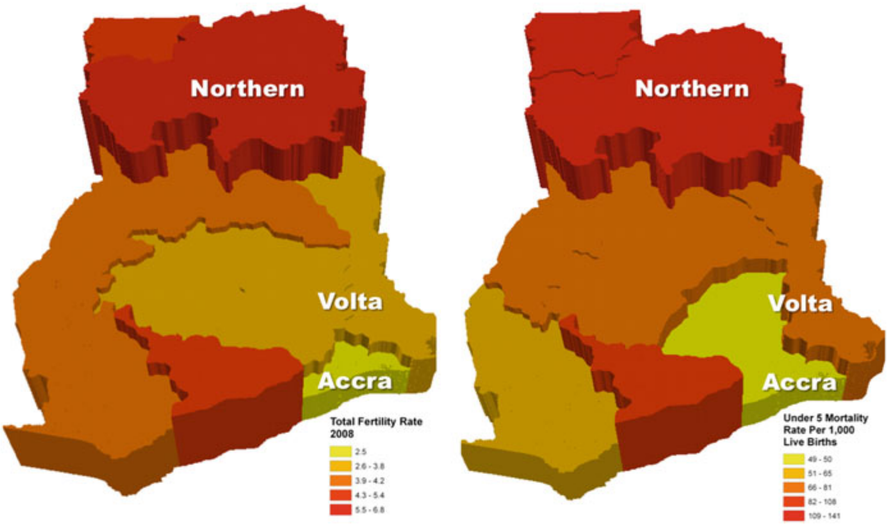

Fig. 6.3 Total fertility rates and under-five mortality rates by region, Ghana, 2008 (Source: Created by the author from Ghana Demographic and Health Survey 2008 data (www.measuredhs.com))

both rates tend to be highest in the north and lowest in the south, with the lowest levels showing up in the Greater Accra Region. This same pattern has prevailed over a long period of time, going back at least to the 1988 GDHS (not shown), but the trends are not the same from one region to another. In 1988, the lowest TFR was in the Greater Accra Region (4.7) and that region was still lowest in 2008, having dropped to 2.5. By contrast, the Northern Region had a TFR of 6.9 in 1988, and that was virtually unchanged (6.8) in 2008. On the other hand, the Volta region had a TFR of 7.2 in 1988 and that had been cut nearly in half, to 3.8, by 2008.

Consistent with demographic transition expectations, there is a high correlation between deaths of children under the age of 5 (U5MR) and fertility. When children have a high probability of dying, the response is almost always high fertility. There is little incentive to limit the number of children being born when death rates are high, but as mortality declines there is almost always a time lag before people realize that the decline in death rates is real, and that coping with more surviving children can be very difficult, eventually leading to a decline in fertility. Greater Accra Region was tied for the lowest U5MR in 2008 at 50 deaths to children under the age of five per 1,000 live births. It was tied with the Volta Region, where fertility has dropped most dramatically, even though for the time-being its TFR is well above that of Accra. At the other extreme, the Northern region had the highest U5MR in 1988 (well above 200), and it still had the highest rate in 2008, even though it had dropped to 140. In other words, the recency and still modest decline in child mortality had not yet influenced fertility in the Northern Region as of 2008. Indeed, it is important to emphasize the recency of mortality declines in Ghana, as elsewhere in West Africa. In 1988, every region in Ghana had an U5MR of at least

100, meaning that 100 out of every 1,000 children were dying before their fifth birthday. As of 2008, there were still three regions (Northern, Central, and Upper West) where U5MR exceeded 100. For comparative purposes, we can note that U5MR in the US is 8 per 1,000, and it is only 3 in Japan.

At each spatial scale the demographic patterns become a bit more complex, emphasizing the point that there are a lot of different ways in which each of the components of demographic transitions can manifest themselves. Although we do not yet have detailed data from the 2010 census, we do have those data for the 2000 census, building on the 10 % sample of individual census returns available through the International Public Use Microdata Sample (IPUMS) project (www.ipums.org) at the Minnesota Population Center (2011), based on data provided by Ghana Statistical Service, from which we also obtained the digital boundary file of the 110 districts (administrative units within Regions) for Ghana.

The census data do not provide the same kind of information about current fertility as are available in surveys such as the GDHS. So, we cannot readily calculate a measure of TFR, but as a proxy we calculate a measure called CEBz, which is the number of children born to date (CEB) for each woman (her parity) as a standard deviate (z) relative to all women in the country of the same age (measured in 5-year age groups), which we label CEBz (Weeks et al. 2013c; Benza 2013):

$$CEBz = \frac{CEB_{individual} - avgCEB_{agegroup}}{Standard\ \ deviation_{agegroup}}$$

The advantage of CEBz is that it provides a single age-adjusted measure (standardized variable) that can be directly correlated with the environmental context measure that is the focus of this research. Unlike the TFR, the CEBz is measured initially at the individual level and then can be readily averaged over any spatial unit. The more traditional way to approach the measurement of fertility at the individual level is to use the number of children ever born alive (CEB) to a woman while controlling for age in the regression model. We have found, however, that the combination of CEB and age is a good predictor of CEBz ($R^2 = .74$), and so we feel that the CEBz measure is a robust index of reproductive levels. It has the disadvantage of not being as directly interpretable as the TFR, but for our spatial analysis we are largely concerned with relative differences between areas. One way to cross-reference this is to note that the average CEBz per region in Ghana based on the data from the 2000 census has a very high correlation ($R = .87$) with the TFR for regions based on the 2003 GDHS.

6.5.3 *Fertility Differences by District Within Ghana*

By averaging the CEBz for women within each of Ghana's 110 Districts, we are able to see a more nuanced spatial pattern of reproductive levels than was apparent in Fig. 6.3. The left side of Fig. 6.4 maps the levels of CEBz per district, with colors

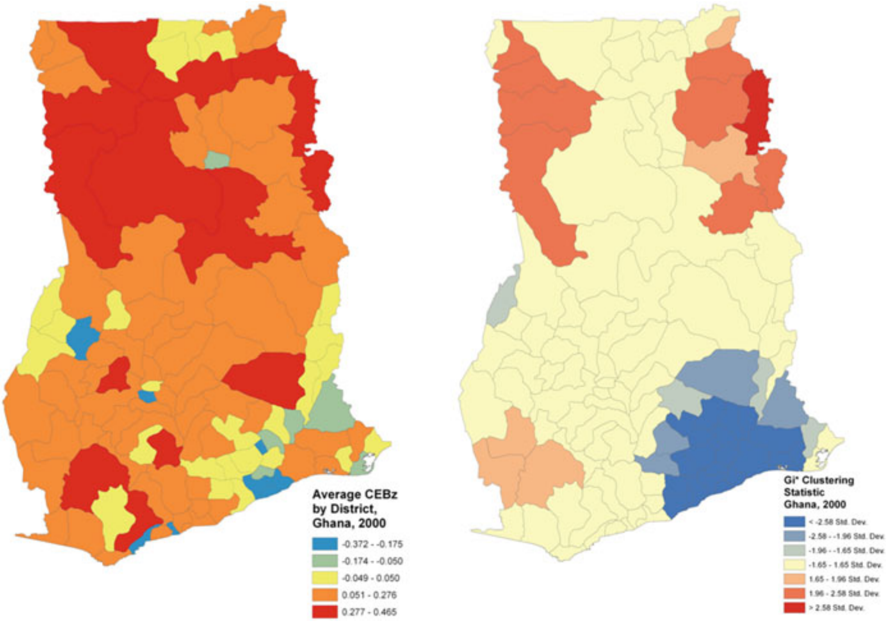

Fig. 6.4 CEBz by district, Ghana, 2000 (Source: Created by the author from 2000 Ghana Census of Population and Housing data (data courtesy of Minnesota Population Center 2011))

in the blue range indicating levels of fertility that are below the national average and colors in the orange/red range reflecting fertility that is above the national average. As was evident in Fig. 6.3, there is a general north-south drift to fertility, with lowest levels in the south and highest levels in the north. The one pocket of below average fertility in the north is the city of Tamale, in the Northern Region. Our first step in quantitatively assessing the spatial pattern of fertility is to test whether or not the observed pattern could have occurred just by chance alone. If not, then we are observing a statistically significant pattern of spatial autocorrelation. As John Logan also notes in his chapter in this volume, the Moran's I statistic is typically used for this task, and in the map in Fig. 6.4 the value of I is 0.33, based on an inverse distance spatial weights matrix using a distance threshold of 90 km, with an associated z-score of 6.99, indicating a highly significant pattern of spatial clustering.

After confirming that there is spatial autocorrelation, the next step is to figure out where the clustering is occurring. The right-side of Fig. 6.4 shows the pattern of spatial clustering, in which districts that are near to other districts with similar levels of fertility are highlighted. In this instance, I have used the Getis-Ord Gi* statistic. Positive values of Gi* that exceed a z-score of 1.96 (the .05 level of statistical significance) indicate spatial association of high values (reds), whereas negative values of Gi* that are less than −1.96 indicate spatial association of low values (blues) (see Mitchell 2005 for an overview of these statistics).

What might explain these spatial patterns? I have already discussed three key interrelated factors that might help us explain why fertility is low (and clustered) in some places while high (and clustered) in other places. The first of these is the decline in mortality, which affects children first and most noticeably, because declining infant and childhood mortality increases the number of surviving children, forcing families and communities to adjust. Thus, we would expect that places with lower infant and childhood mortality will have lower fertility. The second factor is education, going back to the invention of the printing press, which exponentially expanded the diffusion of knowledge, leading to the Enlightenment and all of the scientific discoveries that have come along in the past 200 years, including those that brought about our improved control over mortality. So, everywhere we go in the world, we expect that women who are literate (as a basic measure of education) will have higher levels of survival among her children, and will have adjusted to that by having fewer children (see, for example, Lutz et al. 2010; Courbage and Todd 2011). The third factor is urbanization, since the nature of urban places is that they bring together the elements of change that characterize the modern world.

6.5.4 Data and Methods

The calculation of under-five mortality technically requires a complete reproductive history of women, such as those collected in the Demographic and Health Surveys, and discussed above. The census questionnaire is much less complete, so we cannot be as precise about the level of child mortality when we are using just the census data. However, it turns out that the human regularities associated with patterns of death allow us to make reasonable inferences about this probability from the two questions asked on the census about (1) children ever born to a woman and (2) the number of those children who are still surviving. These two questions do not directly allow us know what fraction of children who were born alive died before reaching age 5, but the work of William Brass (e.g., Brass 1971) and his successors has demonstrated that empirically the ratio of children dead to children born alive to women aged 30–34 is a very close approximation to under-five mortality (Popoff and Judson 2004). In our research focusing on Accra, Ghana, we have found that a simple dichotomous variable of whether or not a woman of reproductive age has ever lost a child who was born alive provides a good individual level indicator of child mortality risks and that, when aggregated, this measure is highly correlated with more sophisticated measures of child mortality (Jankowska et al. 2013).

The census asked people if they could read and write and if so they were deemed to be literate, and so each woman was coded as being literate or not. For aggregation to the districts, however, I used the percent literate among women of reproductive age, since they are the people having the babies that are the focus of the analysis. Almost everywhere in Ghana men are more likely to be literate than are women (there is about a 15 percentage point gap), but the correlation between male and female literacy levels is very high (R = .96).

The urban variable was based on coding whether or not the person was living in an area that was urban according to the definition of Ghana Statistical Service. This means that the place must have at least 5,000 people. Thus, each person was coded as being simply urban or rural, but at the district level the variable becomes the percentage of the total population that is living in an urban place.

The results of an ordinary least squares (OLS) regression are that the three predictor variables—percent of women of reproductive age with at least one live birth who have lost at least one child (proxy for under-five mortality), the percent of women of reproductive age who are literate (proxy for education), and the percent of the population living in an urban place—are able to explain 74 % of the variation from district to district in CEBz (our measure of fertility). The most important predictor was percent urban (standardized beta coefficient of −.476), the next most important was female literacy (beta coefficient = −.305), with child mortality being close behind at .256.

In classic statistical thinking, this might be the end of the story. The three key variables that were identified as potential predictors of fertility are, in fact, empirically linked to fertility in the predicted direction. But classical statistics assumes that the residuals are spatially random, whereas in the real world things tend to be spatially clustered, even when they haven't been measured in the regression equation. So, the next step in any regression model using georeferenced data is to test the residuals for spatial autocorrelation. As already noted, this is accomplished with the global Moran's I statistic, which in this case is 0.18, with an associated z-score of 3.95, indicating a statistically significant level of clustering in the residuals. A map of the residuals (not shown) indicates that fertility levels are lower than predicted in the south of the country, while being higher than predicted in the north. The overall importance of this pattern can be evaluated by running a spatial regression model, using GeoDa (http://geodacenter.asu.edu/) or comparable software. In GeoDa, the diagnostics for spatial dependence in the data for Ghana districts indicate that one or more spatially correlated covariates has been omitted from the model, thus influencing our overall inferences about the predictors of fertility. The variable(s) are unknown, but it is known that it/they are spatial correlated. It will take some scientific detective work to figure out what the variable(s) might be.

In response to this information, a spatial error autoregressive model was run, which improved the overall R-squared slightly from .74 to .78, but the spatial variable (a proxy for the missing variable(s) in the analysis) was still less important statistically than the other three predictor variables. The map of residuals suggested that the spatial variable(s) reflected some kind of cultural difference between the north and the south of the country. So, a dummy variable was constructed in which districts in the three northern regions were coded as zero and districts in the seven southern regions were coded as one. When this variable was added to the regression model, the results were virtually identical to the spatial error model—an R-square of .78 and the standardized residuals were no longer statistically significant. This indicated that the "missing" spatial component had been captured, although it will take further investigation to know exactly what it is about the culture of the northern areas that sets them apart from the southern regions. Religion and ethnicity probably play important roles.

Although the "mystery" of the spatially autocorrelated regression residuals was solved, there is more to the spatial diagnostics. We know that fertility is spatially clustered (see Fig. 6.4), and that the combination of child mortality, female literacy, urbanization, and being in the north or south regions of the countries explains more than three-fourths of the variation in fertility, but which of these variables helps us to explain the spatial patterns that we see? One way to approach this is to decompose each of the predictor variables into their spatial and non-spatial components, in a process known as spatial filtering (Getis 1995; Getis and Griffith 2002; Weeks et al. 2004; Griffith 2010). In this statistical approach, we first test for the presence of spatial dependence in each of the predictor variables by calculating Moran's I, using an inverse of squared distance weights matrix, where distance is measured between the centroids of districts. For each spatially dependent independent variable, we use the Gi(d) statistic as a spatial filter to extract the spatially autocorrelated portion of that variable. The difference between the original variable x(i) and the filtered variable x(f) is a new variable x(sp), that represents the spatial effects embedded in x(i). These two variables, x(f) and x(sp) replace the original variable x(i) in the regression equation to produce a spatially filtered regression model in which the contribution of the spatial and filtered (nonspatial) components of each variable can be determined by the beta coefficients in the resulting model. This technique of spatial filtering has been programmed in Fortran by Scott (1999), and is available from the author by request as an ArcGIS toolbox.

Female literacy is the most spatially clustered of the three major predictor variables (Moran's I (z) = 10.2), followed by child mortality (Moran's I (z) = 5.6), whereas the percent urban was not spatially clustered (Moran's I (z) = 0.0). The regional variable is obviously clustered and was not involved in spatial filtering. Since the percent urban is not clustered, it was not filtered, whereas both female literacy and child mortality were spatially clustered and so they were filtered. The new regression model has the same R-square (.78), but a more nuanced interpretation of the predictor variables, as shown in Table 6.1. The filtered (non-spatial) component of female literacy becomes the most important predictor of fertility, with a standardized beta coefficient of −0.447, indicating that as literacy goes up, fertility goes down, regardless of where you are. There is also a spatial component of literacy, also negative, indicating that the spatial clustering of female literacy is also important, at least partly because female literacy is highest in the south of the country. Percent urban is also negatively associated with fertility, as expected. Child mortality is positively associated with fertility, as expected, and the filtered component is more important than the spatial component, which is just on the edge of being statistically significant. Finally, it can be seen that the dummy variable for being in the north or south of the country is statistically significant, even after accounting for the fact that the north is less urban and has lower levels of literacy. This may be due to the fact that the decline in child mortality is more recent

Table 6.1 Spatially filtered regression model results

Predictor variable	Standardized beta coefficient	t-score	p-value	VIF
Lostchild filtered	0.243	4.595	0.000	1.314
Lostchild spatial	0.173	2.023	0.046	3.416
Female literacy filtered	−0.447	−5.784	0.000	2.804
Female literacy spatial	−0.280	−3.783	0.000	2.574
Being in southern regions	0.343	3.943	0.000	3.560
Percent urban	−0.421	−7.065	0.000	1.665

Dependent variable is CEBz
R-square = .78

in the north than in the south, and so fertility remains high because there has not been enough time for couples to react to the increased probabilities of children surviving. We know that migration out of the north is an important demographic component of change, as larger families cannot subsist on the same land. Since job opportunities are greatest in the cities of the south, young migrants tend to head that direction, thereby exposing them to the different demographic regime of the south. Since there is a fair amount of return or circular migration within the country, this will almost certainly have the effect over time of diffusing new ideas to the north of the country.

The fact that there are clear differences between the north and the south in Ghana alerts us to test for spatial heterogeneity, the situation in which the regression coefficients vary from one place to another, rather than remaining the same everywhere, which is the assumption underlying classical regression analysis. However, it turns out that the results from a geographically weighted regression (GWR) implemented in ArcMap 10.1 do not suggest the presence of spatial heterogeneity. The overall pseudo R-square from GWR is virtually identical to the ordinary least squares results shown in Table 6.1. Although the coefficients show a north-south drift, the differences are not statistically significant.

A key component of spatial demography is the emphasis on scale, as discussed above. Relationships may be different depending upon the scale of the data with which you are working. This is not unlike the concept of spatial heterogeneity, except that it might be called scalar heterogeneity. It is not that coefficients differ according to where you are, but rather that they differ according to the scale of the data. This is sometimes subsumed under the category of the "ecological fallacy," which refers to the idea that relationships found for aggregated data may not hold at the individual level. In general, the variability at the individual level is much greater than the variability observed in the aggregate, and different levels of aggregation may produce different levels of variability. There is not one "true" answer here. Rather, data need to match the research questions and the interpretation of data needs to match the results.

If we compare individual level data from the 2000 Ghana census with the district level analysis that I have just discussed, we find that, as is typically the case, the relationships are less robust. We are able to explain only 10 % of the variability from woman to woman in CEBz on the basis of her literacy, experience losing a child, urban residence, and living in the south compared to the north. Moreover, at the individual level, having lost a child is the single most important predictor of fertility levels, followed by living in an urban rather than a rural environment, then female literacy, and finally living in the south rather than the north. So, the individual level analysis is quite different than the analysis aggregated at the district level. Again, it is not that one set of relationships is wrong and the other right, but rather that each analysis answers a different research question and leads to different policy implications. The "fallacy" would be to believe that the analysis at the district level applied to individual women. It clearly does not. On the other hand, the spatial relationships in the aggregate tell us a lot about the overall trends in fertility and its correlates, even if they tell us less about what is happening to a particular woman in any one of those areas. At the same, narrowing in on a specific area can provide important local knowledge that gets lost in the bigger picture.

6.5.5 Fertility in the Context of the District That Is the Capital City Accra

In the Greater Accra Region there were five districts as of the 2000 census and one of those encompasses the Accra Metropolitan Area (AMA), which has been the site of intensive analysis by my research colleagues and I (Weeks et al. 2013b; Engstrom et al. 2013; Verutes et al. 2012). We have documented that there is considerable variability in fertility within the city, ranging from fertility levels comparable to Saudi Arabia (e.g., high) to those comparable to Belgium (e.g., low). But, as we note (Weeks et al. 2013a:16), the variability is compounded by spatial complexity:

> . . .the processes of social sorting operate in powerful but distinctive ways in Accra and possibly in other African cities. In Accra, rich and poor live closer to one another than in many European or North American cities, following patterns which are more reminiscent of the living conditions in 19th-century industrial cities (Booth 1969 [1902]). In addition to patterns of compound living which persist in the city to this day, the lack of effective planning controls means that squatting in temporary housing, so-called kiosks or containers is quite common throughout the urban area, including even in the better off neighbourhoods. There are clearly different patterns of social identity in African cities compared with elsewhere. Race and color may not have the same meaning as elsewhere, but certainly the census data indicate persistent and strong preferences for intermarriage within the same ethnic or language groups (Weeks et al. 2011). Pentecostal and charismatic churches have recruited members from a wide range of social strata, adding to the complex mix of people who worship and socialise together (Gifford 2004). Further, the high levels of literacy in the population means that health and other messages are widely received through FM radios, televisions and, increasingly through social media, thus breaking down some of the barriers between the less-educated and better-educated that are found in many other urban environments.

That complexity is not easy to capture with a single regression equation. Indeed, you can appreciate that we cannot completely replicate the district level analysis discussed above at the finer spatial scale of the city of Accra. All women live in the south and all women live in an urban place, so those two variables are now constants. We are left to predict CEBz with literacy and child mortality, recognizing that most women in Accra are literate and child mortality is lower in Accra than most other places in the country. At the individual level within Accra, those two variables explain only 6 % of the variation from woman to woman in her fertility relative to women her age in the entire country. However, as was true for the entire country, her experience with having lost a child is most closely associated with fertility, with literacy being less important. Note that if we compare women in Accra only with themselves in terms of fertility, rather than with women all over the country, we capture more of the variation that exists within the city and at the individual level the R-square increases from .06 to .17, with a woman's experience of losing a child still being more important than literacy as a factor predicting her fertility.

Another complexity within Accra is that a powerful predictor of fertility at the neighborhood level is whether or not women have ever been married. There are relatively few out-of-wedlock births in Accra, and delaying marriage has become a key element in low fertility within the city (Weeks et al. 2010). Within the district of the AMA, there are 1,723 enumeration areas (EAs), similar to census blocks in the US, and they are the smallest unit of census geography. When we aggregate individual level data from the census to the EA level, we find that the proportion of women of reproductive age who have lost a child, along with the proportion of women who are literate, explains 39 % of the variation from EA to EA in fertility. When we add in the proportion of women of reproductive age who have never married, the R-square jumps to 50 % and the never-married variable is the single most important predictor. Women born in Accra are more likely to be single than those born outside the city, are more likely to be of the Akan ethnic group than of the Ga ethnic group (the two largest ethnic groups in the city), are more likely to have completed a secondary (high school) level of education, but are less likely to be working than women who are married, and those who do work tend to be domestic workers, apprentices, or unpaid family workers, and they are more likely to live in higher status neighborhoods. They are in some ways more "western" or "modern," suggestive of the cultural component of delaying marriage and childbearing that we associate especially with the urban transition.

6.6 Conclusion

What can be learned from the spatial analysis of demographic data that would not have otherwise been gleaned from the non-spatial "classical" approach? The single most important thing, in my opinion, is that spatial demography alerts us to the complexity of the world in which we live. The several suites that comprise the overall demographic transition provide us with a framework for understanding the world demographically, but we cannot be so smug as to believe that anywhere we go in the world, the exact same kinds of relationships will hold true. Cultural variability is too strong to allow that to happen and culture tends to have a strong spatial component. Spatial analyses in demography have allowed us to use a whole new set of "microscopes" to improve our knowledge of the world. Just as new technology has benefited astronomy and biology and many other sciences, the application of the emerging fields of spatial science and spatial statistics to the science of human populations allows us to see, and thus try to explain, variations in human demographic behavior that were previously hidden from us. Just as cancer was once thought of as a single type of disease, scientific technology has allowed the identification of many different types of cancers, permitting researchers and practitioners to improve levels of prevention, treatment, and thus to extend human longevity. In demography, the new spatial technologies promise similarly to allow us to more precisely identify the trends taking place throughout the different aspects of the demographic transition, with the goal of expanding our capacity to understand and improve the human condition.

Acknowledgements This research was funded in part by grant number R01 HD054906 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development ("Health, Poverty and Place in Accra, Ghana," John R. Weeks, Project Director/Principal Investigator). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute of Child Health and Human Development or the National Institutes of Health.

References

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Benza, M. (2013). Living arrangements and fertility: A case study in southern Ghana. In J. R. Weeks, A. G. Hill, & J. Stoler (Eds.), *Spatial inequalities: Health poverty and place in Accra, Ghana*. Dordrecht: Springer Publishing Co.

Booth, C. (1969 [1902]). *Life and labour of the people of London, 2nd Series*. New York: A.M. Kelley.

Boserup, E. (1981). *Population and technological change: A study of long-term trends*. Chicago: University of Chicago Press.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), *Biological aspects of demography*. New York: Barnes and Noble, Inc.

Chou, Y.-H. (1997). *Exploring spatial analysis in geographic information systems*. Santa Fe: OnWard Press.

Coale, A. (1973). The demographic transition. In IUSSP (Ed.), *Proceedings of international population conference* (Vol. 1, pp. 53–72). Liege: IUSSP.

Courbage, Y., & Todd, E. (2011). *A convergence of civilizations: The transformation of Muslim societies around the world*. New York: Columbia University Press.

Cressie, N. A. C. (1993). *Statistics for spatial data* (Rev. ed.). New York: Wiley.

Davis, K. (1949). *Human society*. New York: Macmillan.

Davis, K. (1963). The theory of change and response in modern demographic history. *Population Index, 29*(4), 345–66.

Davis, M. (2007). *Planet of slums*. London: Verso.

Demeny, P. (1968). Early fertility decline in Austria-Hungary: A lesson in demographic transition. *Daedalus, 97*(2), 502–22.

Easterlin, R. (1978). The economics and sociology of fertility: A synthesis. In C. Tilly (Ed.), *Historical studies of changing fertility*. Princeton: Princeton University Press.

Engstrom, R., Rain, D., Ofiesh, C., Jewell, H., & Weeks, J. R. (2013). Defining neighborhood boundaries for urban health research in developing countries: A case study of Accra, Ghana. *Journal of Maps, 9*(1), 36–42.

Ferguson, N. (2011). *Civilization: The west and the rest*. New York: Penguin.

Fotheringham, A. S., Charlton, M., & Brundson, C. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.

Getis, A. (1995). Spatial filtering in a regression framework: Examples using data on urban crime, regional inequality, and government expenditures. In L. Anselin & R. Florax (Eds.), *New directions in spatial econometrics* (pp. 172–185). Berlin: Springer.

Getis, A., & Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical Analysis, 34*(2), 130–140.

Ghana Statistical Service. (2012). *2010 population & housing census: Summary report of final results*. Accra: Ghana Statistical Service.

Gifford, P. (2004). *Ghana's New Christianity*. Bloomington/Indianapolis: Indiana University Press.

Griffith, D. A. (2010). Spatial filtering. In M. M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis*. Heidelberg: Springer.

Jankowska, M., Benza, M., & Weeks, J. R. (2013). Estimating spatial inequalities of urban child mortality. *Demographic Research, 28*(2), 33–62.

Leasure, J. W. (1982). L' Baisse De La Fecondité Aux États-Unis De 1800 a 1860. *Population, 3*, 607–22.

Lesthaeghe, R. J. (1977). *The decline of Belgian fertility, 1800–1970*. Princeton: Princeton University Press.

Lutz, W., Crespo Cuaresma, J., & Abbasi-Shavazi, M. J. (2010). Demography, education, and democracy, global trends and the case of Iran. *Population and Development Review, 36*(2), 253–281.

Minnesota Population Center. (2011). *Integrated public use microdata series, international: Version 6.1 (Machine-readable database)*. Minneapolis: University of Minnesota.

Mitchell, A. (2005). *The Esri Guide to Gis Analysis, volume 2: Spatial measurements & statistics*. Redlands: ESRI Press.

Norris, P., & Inglehart, R. (2004). *Sacred and secular: Religion and politics worldwide*. New York: Cambridge University Press.

Popoff, C., & Judson, D. H. (2004). Some methods of estimation for statistically underdeveloped areas. In J. S. Siegel & D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed.). San Diego: Elsevier Academic Press.

Ravenstein, E. G. (1876). *The birthplaces of the people and the laws of migration*. London: Trübner and Company.

Scott, L. (1999). *The accessible city: Employment opportunities in time and space*. Ph.D. Dissertation, Department of Geography, San Diego State University, San Diego.

Star, J., & Estes, J. (1990). *Geographic information systems: An introduction*. Englewood Cliffs: Prentice Hall.

Teitelbaum, M. (1975). Relevance of demographic transition for developing countries. *Science, 188*, 420–425.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 26*, 234–40.

Tobler, W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers, 94*(2), 304–310.

United Nations Population Division. (2011). *World population prospects: The 2010 revision*. New York: United Nations.

Verutes, G., Benza, M., Coulter, L., & Weeks, J. R. (2012). Health, poverty and place in Accra, Ghana: Mapping neighborhoods. *Journal of Maps, 8*(4), 369–373.

Voss, P. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*, 457–476.

Weeks, J. R. (2004). The role of spatial analysis in demographic research. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science: Examples in best practice*. New York: Oxford University Press.

Weeks, J. (2012). *Population: An introduction to concepts and issues* (11th ed.). Belmont: Wadsworth Cengage Learning.

Weeks, J. R., Getis, A., Hill, A. G., Gadalla, M. S., & Rashed, T. (2004). The fertility transition in Egypt: Intra-urban patterns in Cairo. *Annals of the Association of American Geographers, 94* (1), 74–93.

Weeks, J. R., Getis, A., Hill, A. G., Agyei-Mensah, S., & Rain, D. (2010). Neighborhoods and fertility in Accra, Ghana: An amoeba-based approach. *Annals of the Association of American Geographers, 100*(3), 558–578 PMCID: PMC3093308.

Weeks, J. R., Agyei-Mensah, S., Owusu, G., Hill, A. G., & Benza Fiocco, M. (2011). Ethnic assimilation in Accra, Ghana. In *Annual meeting of the Population Association of America*. Washington, DC.

Weeks, J. R., Hill, A. G., & Stoler, J. (2013a). An introduction to the "Accra School" of spatial inequalities and demography. In J. R. Weeks, A. G. Hill, & J. Stoler (Eds.), *Spatial inequalities: Health poverty and place in Accra, Ghana*. Dordrecht: Springer.

Weeks, J. R., Hill, A. G., & Stoler, J. (Eds.). (2013b). *Spatial inequalities: Health poverty and place in Accra, Ghana*. Dordrecht: Springer.

Weeks, J. R., Stoler, J., Hill, A. G., & Zvoleff, A. (2013c). Fertility in context: Exploring egocentric neighborhoods in Accra. In J. R. Weeks, A. G. Hill, & J. Stoler (Eds.), *Spatial inequalities: Health, poverty, and place in Accra, Ghana*. Dordrecht: Springer.

Part II
Research Practice in Spatial Demography

Chapter 7
Modeling 'Dependence of Relevant Alternatives' in Consumer Choice: A Synthesis from Disparate Literatures

Lee Rivers Mobley and Gloria J. Bazzoli

7.1 Introduction

In this paper, we are concerned with modeling consumer choice among competing products when spatial location matters to consumers as a product attribute. This is certainly the case in the consumption of medical services. We summarize and synthesize several areas of research from disparate literatures, to show that incorporating spatial dimensions can enhance model tractability and plausibility. We draw from the hospital choice literature, the regional science literature on spatial modeling, and the new empirical industrial organization literature on estimating choice models in differentiated product industries.

Many studies of hospital choice by consumers have been based on McFadden's conditional logit model (1974), which has the well-known 'independence of irrelevant alternatives' (IIA) property (Burns and Wholey 1992). What this IIA property means is that predicted choice probabilities are independent of the size and composition of the choice set, and thus do not adequately represent product substitution effects. For example, Feldman et al. (1989) argue that an IIA model is inappropriate for modeling choice of health plan, because the addition of a new plan is more likely to affect the choice of close substitutes than far. The relative values of predicted probabilities of available choices should (realistically) change to reflect any change in the choices and consequent substitution patterns among them. This

L.R. Mobley, Ph.D. (✉)
School of Public Health and Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA 30303, USA
e-mail: lmobley@gsu.edu

G.J. Bazzoli, Ph.D.
Department of Health Administration, Virginia Commonwealth University, 980203, Richmond, VA 23298-0203, USA
e-mail: gbazzoli@vcu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_7

invariance under IIA leads to unrealistic parameter estimates that should not be used in policy conclusions regarding the value of these changes to consumers, as we discuss further below.

This IIA property manifests itself when the odds ratio of choices among two alternatives is invariant to the introduction of a third alternative – an assumption which may be viable for some products and services, but seems unlikely for hospital choice. In this paper a different approach to modeling consumer choice is described, drawing from the spatial modeling literature, which is compared to some recent methods developed in new empirical industrial organization (NEIO) models of differentiated-product demand systems. It is demonstrated that, when product location matters to consumers, the inclusion of ‘accessibility factors’ in the choice model addresses the ‘independence of irrelevant alternatives’ property. This alternative model, which we classify as a ‘dependence of relevant alternatives’ (DRA) model, has a tractable Type I Extreme Value error distribution, and is appropriate for modeling healthcare provider choice in modern urban markets.

We review the literature on hospital choice and find many older studies using IIA models, and thatnewer, more sophisticated models exploit what we call here a ‘dependence of relevant alternatives’ (DRA) formulation. Perhaps the observed scarcity in supply of the DRA-type models in hospital choice and related research until very recently was due to the difficulty inherent in estimating the increasingly more sophisticated alternatives (Jones 2000). Perhaps some researchers in the past have discovered what we do here – that a tractable model exists for situations where location matters – but due to the complexities once inherent in creating the necessary spatially-referenced variables, this avenue has not been fully exploited. However, with improvements in GIS and spatial modeling software, these things are no longer so difficult, as described herein. We hope that in explicitly comparing some IIA and DRA models, and describing some recent examples of excellent work of the latter variety, that the value of a spatial approach to healthcare choice problems will be highlighted and embraced more broadly in health economics research.

7.2 Destination Choice Models: Some Comparisons

Various models have been proposed by economists and spatial scientists to characterize the choice by consumers over spatially-distributed alternatives. What follows draws from Fotheringham (1983) and Fotheringham and O’Kelly (1989), who demonstrate the similarities and differences among various approaches.

The conditional logit model of McFadden (1974) is a random-utility framework which describes consumer i’s choice of a particular product j, based on product j’s characteristics as perceived by consumer i (X_{ij}) and individual i’s characteristics (Z_i). In spatial terms the analogous model has an individual at origin i choosing destination j, where product j’s characteristics (as perceived by consumer i, X_{ij}) include the distance between origin and destination (d_{ij}). The conditional logit model of McFadden (1974) and the production-constrained gravity model are essentially equivalent, when distance between origin and destination (d_{ij}) is included as an explanatory variable in the logit model. Both have a basis in the random utility maximization model (Fotheringham and O'Kelly (1989), p. 73, McFadden (1981)).

The Random Utility Function Model:

$$U_{ij} = \beta X_{ij} + \gamma_j Z_i + \varepsilon_{ij} \tag{7.1}$$

where

$V_{ij} = \beta X_{ij} + \gamma_i Z_i$, the stochastic/deterministic portion of random utility
ε_{ij} = the error process which follows a Type I Extreme Value distribution
j = 1, 2,.. J products/destinations
i = 1, 2,. . . I consumer individuals/origins
X_{ij} are the vectors of values of the attributes of the jth choice as perceived by the ith individual, which can include the distances between origin (individual) i and destination(choice) j: d_{ij}
Z_i are characteristics of individuals at origin i

McFadden's conditional (multinomial) logit model (1974) is described fully in Maddala (1983, pp. 59–61) and in Fotheringham and O'Kelly (1989, pp. 73–74). The additive error term allows one to assume that errors are distributed i.i.d. with a Type I extreme value distribution, so the model is empirically tractable. The probability that destination j is chosen over other j' alternative destinations by consumer i, is:

$$Pij = \frac{\exp(Vij)}{\sum_{j \in Z} \exp(Vij)} \tag{7.2}$$

whereV_{ij} is defined above. Because of the i.i.d. error assumption, this model has the property referred to in the literature as IIA (independence of irrelevant alternatives). Thus, the odds ratio for the jth and j′th choices is the same irrespective of the total number of choices considered. That is, the addition of another alternative k does not affect the odds ratio of alternatives j and j′ (as these are reduced proportionately):

$$\frac{Pij}{Pij'} = \frac{\exp\left(\beta Xij + \gamma Zi\right)}{\exp\left(\beta Xij' + \gamma Zi\right)} \tag{7.2′}$$

Both McFadden's conditional logit and the production-constrained gravity models exhibit the IIA property (Fotheringham and O'Kelly 1989, pp. 73–74). This restriction may be useful in some applications, such as in transportation planning (McFadden and Reis 1975), but it is clearly inappropriate in others. In particular, when the conditional logit consumer choice model is used as the basis for a demand system, this IIA property manifests as restrictions on the own- and cross-price elasticities and product substitution patterns across consumers (McFadden (1981); Berry et al. (1995); Nevo (2000, 2001)). For example, the cross-price elasticitiesmay have functional dependence on market shares only, thus two products with equal market shares will have equal substitution effects with a third product, irrespective of other product attributes (Nevo 2001). Nevo uses as an example the change in price of one type of children's cereal leading consumers to substitute towards other brands in proportion to their market shares, irrespective of whether they are children's cereal or health-conscious cereal types.

Implausible substitution patterns are implied by the conditional logit model because the slope of demand in both own and rivals' prices are dependent only upon choice probabilities, i.e. two products with identical choice probabilities will have identical own-price and cross-price slopes with other products. If consumers are identical, these undesirable properties of individual level demand curves carry over to the aggregate market level demand curves faced by firms. But if consumers are heterogeneous and there are interactions between consumers and the firm characteristics, then these undesirable characteristics need not carry over to market demand curves. This would mean relaxation of the restrictive i.i.d. assumption for the error term. Replacing the i.i.d. assumption with a variance components structure leads to less restrictive models (Nevo 2001, p. 12). These include the Generalized Extreme Value Model (McFadden 1978), the Nested Logit Model (McFadden 1981) and the Principles of Differentiation Generalized Extreme Value Model (Bresnahan et al. 1997). While less restrictive, these models derive substitution patterns from *a priori* segmentation of choice clusters, which is described next for the Nested Logit Model. This *a priori* segmentation poses problems for spatial choice models, which we explore in turn.

7.3 The Nested Multinomial Logit Model

The nested multinomial model described by McFadden (1981) is not characterized by the IIA property, and is seen as a viable alternative to the conditional logit model when the choice process is hierarchical. This model is appropriate when the form of the hierarchy is known to the modeler, because it involves identifying a subset of all alternatives as the choice cluster for each individual. The error term follows a Generalized Extreme Value distribution, which necessitates the *a priori* identification of a set of clusters of alternatives. Hierarchical decisions are made when individuals first compare all clusters, and then select one particular cluster within which to evaluate individual alternatives. Maddala (1983, pp. 67–70) describes this model; Fotheringham and O'Kelly do also (1989, pp. 77–78):

The probability that individual i will select a particular cluster s′ is:

$$P_{is'} = \frac{\exp(V_{is'})[\sum_{j\in s'} \exp(V_{ij})]^{\sigma}}{\sum_s \exp(V_{is})[\sum_{j\in s} \exp(V_{ij})]^{\sigma}} \tag{7.3}$$

The probability that individual i will select a particular alternative j′ within the chosen cluster s′ is

$$P_{ij'\in s'} = \frac{\exp\left(V_{ij'\in s'}\right)}{\sum\limits_{j\in s'} \exp\left(V_{ij}\right)} \quad \text{for } j' \in s' \tag{7.4}$$

Thus the joint probability that individual iselects j′ from the set of all alternatives is the product of the conditional and marginal probabilities above:

$$P_{ij'} = P_{is'} \ * \ P_{ij'\in s'} \qquad \text{or} \qquad P_{ij'} = (\text{eqn } 7.3) * (\text{eqn } 7.4)$$

Because choice of j by consumer i now depends upon j being included in the initial cluster s′, the introduction of a spatially proximate (*substitute*) choice alternative k can affect the probability of j's inclusion in the cluster, hence the odds ratio (Pij/Pij′) can now change with the introduction of another alternative. Thus the nested logit model is characterized by '*dependence* of relevant alternatives' (DRA).

Following the literature, we define the 'inclusive value' (in numerator brackets, Eq. 7.3):

$$\sum_{j\in s'} \exp\left(V_{ij}\right)$$

This term describes the attractiveness of a cluster, which results from the individual alternatives contained within it. In the nested logit model, the inclusive value is parameterized by σ, which reflects the extent to which individuals process information hierarchically (i.e., the extent to which they first carve out a subset cluster from all available alternatives). When the inclusive value parameter $\sigma = 0$, the nested logit reduces to the conditional logit model, with i.i.d. errors, and is subject to the IIA property. However, when $\sigma \neq 0$, the IIA property no longer holds; error terms are correlated across choices and adding more choices will affect the odds of choice among alternatives.

One problem with applying the nested logit model to spatial choice is that spatial clusters are not easy to identify, and may vary across individuals (Fotheringham and O'Kelly 1989, p. 78). Also, space is continuous, which poses problems when using this discrete choice model – spatial clusters are likely to be 'fuzzy'. When clusters are not known *a priori*, bootstrap methods can be used on a sample subset to calibrate the nested logit model. This essentially replaces the inclusive value term with an observed proportion, which is interpreted as a probability. This practice allows a straightforward analogy between the nested logit and the competing destinations model.

7.4 Competing Destinations Model

The competing destinations model is a logit formulation that accounts simultaneously for both substitution and spatial structure effects. In this model, each alternative's utility is weighted by the *probability* of that alternative being evaluated. The errors are independently and identically distributed with a Type I Extreme Value distribution (which does *not* require *a priori* definition of clusters of alternatives). Both the conditional and nested logit models can be seen as special cases of this model.[1] The choice model has the form:

$$P_{ij'} = \frac{\exp\left(V_{ij'}\right)^{*}L_i\left(j' \in m\right)}{\sum_j \exp\left(V_{ij}\right)^{*}L_i\left(j \in m\right)} \tag{7.5}$$

where m is a subset of the set N of all alternatives, and L_i $(j \in m)$ is the likelihood that individual i perceives alternative j to be in set m. This model does not exhibit the IIA property, because when new alternatives are introduced, they can affect the likelihood that an alternative j is in set m. Thus the odds of choice are sensitive to changes in relative attractiveness of products with new entry, through the Likelihood term:

$$\frac{Pij}{Pij'} = \frac{\exp(\beta Xij + \gamma Zi)^{*}Li(j \in m)}{\exp\left(\beta Xij' + \gamma Zi\right)^{*}Li\left(j' \in m\right)} \tag{7.6}$$

Defining the Likelihood has traditionally proceeded along two dimensions (Fotheringham and O'Kelly 1989). One is to define the likelihood that a particular alternative is in the choice set as a function of its dissimilarity with others in product *attribute* dimensions (Meyer and Eagle (1982), Borgers and Timmermans (1987, p. 14–15). This is accomplished by introducing a term (for the Likelihood) which measures the average degree of dissimilarity between each particular alternative and all others. This Likelihood term, the dissimilarity measure, allows for substitution effects to be present. This approachensures that alternatives with similar

[1] If an individual evaluates all alternatives, this is a conditional logit model; if an individual processes information hierarchically and choice set membership is known, this is a nested logit model.

deterministic utility components and similar dissimilarity measures will have similar choice probabilities (Borgers and Timmermans 1987).

The other approach is spatial, and recognizes that the more proximal are alternatives geographically, the more likely they are to be considered substitutes, which affects the likelihood of inclusion in the choice set. While many different formulations of the spatial competing-destinations model are possible, Fotheringham (1983) suggests defining a 'potential accessibility' measure A_{ij} based on distance as:

$$A_{ij} = \frac{\sum_{j \neq j'} w_j * d_{jj'}}{n-1}, \tag{7.7}$$

where w_j represents the weight of alternative j and $d_{jj'}$ represents the distance between all pairs of alternatives j and j′. Most importantly, the $d_{jj'}$ term models *spatial structure among competing alternatives*, not spatial structure (distance d_{ij}) between origin and destination. Besides accounting for substitution effects among alternatives, these models can simultaneously account for spatial structure effects like spatial competition or spatial agglomeration. To see this, we can follow Fotheringham (1983), and define the Likelihood component as a function of the accessibility measure:

$$\mathrm{L_i}(\mathrm{j} \in \mathrm{m}) = \mathrm{A}_{\mathrm{ij}}^{\theta} \tag{7.8}$$

Substituting (7.8) into (7.5), the competing destination model is:

$$P_{ij'} = \frac{\exp\left(V_{ij'}\right) * \left(A_{ij'}\right)^{\theta}}{\sum_j \exp\left(V_{ij}\right) * \left(A_{ij}\right)^{\theta}} \tag{7.9}$$

The parameter θ reflects the degree to which information is processed hierarchically (analogous to the parameter σ on the inclusive value term in the nested logit model).[2]

The expected sign of θ can be positive or negative, depending upon the empirical context in which the choice model is defined, and on the intrinsic attractiveness of large clusters of alternatives (Fotheringham 1986). If the attraction of a cluster increases exponentially as the number of alternatives within it increase, θ will be positive reflecting some sort of agglomeration or local spillovers from close proximity (shopping mall). If the attraction of a cluster increases logarithmically

[2] It can also reflect the degree of substitutability among competing alternatives (Borgers and Timmermans 1987) when the parameter is allowed to vary across k attributes (θ_k). In this case, if the sum of the parameters θ_k is zero, then there is no product substitution, and the competing destinations model reduces to the conditional logit model.

with its size, θ will be negative reflecting some sort of competition or congestion effects, wherein alternatives in closer proximity to others are less likely to be chosen (residential neighborhoods).

7.5 NEIO Choice Models with Product Differentiation

Some recent additions to the new empirical IO literature (NEIO) focus on estimating demand systems for heterogeneous products. In these works, careful estimation of substitution effects is important, because a primary concern in them is to measure market power or pricing conduct. Thus the traditional conditional logit models do not provide an adequate foundation for the demand systems. A body of literature has evolved which recognizes that unobserved product heterogeneity characteristics (such as unobserved consumers' valuation of product quality) can result in violation of the i.i.d. assumption for the error term in random utility (conditional logit) models (Berry (1994); Berry et al. (1995); Nevo (2001)). These models incorporate components in error terms to reflect unobserved consumer tastes, and this eliminates the problem of *a priori* unreasonable substitution effects (Berry 1994, p. 246). For example, Nevo (2001) incorporates a brand-specific fixed effect for each product in the error term, which reflects unobserved product heterogeneity (quality). Because the new compound error term is no longer independent of product characteristics, the cross-price substitution patterns are allowed to be driven by these characteristics. This is accomplished via the expanded error term without the need for *a priori* segmentation of the market (which we saw above is a complicating factor in the nested logit-type models). Now, an increase in the price of product j affects consumers with different tastes differently – some will substitute toward a particular group of products that closely resemble product j. (Thus we no longer have the problem noted earlier of indifference between a children's cereal and a nutritious cereal with equal market shares).

A problem with this error components approach is empirical tractability for large problems including many consumers and choices. A solution to this problem has been found in recent hospital choice literature, discussed further below.

7.6 Hospital Choice Models

Early hospital choice models used variations of the gravity model to examine the unconditional or simple relationship between distance and hospital utilization (McGuirk and Porell 1984). Conditional choice models have been used more recently, which explicitly incorporate other influences besides distance on the probability that a particular hospital is chosen (Burns and Wholey 1992). Burns and Wholey (1992) introduce physician characteristics in their model, which is of the traditional conditional logit type, subject to IIA. They describe a weak test for the IIA property by Hausman and McFadden (1984), and explain why it may not be reliable (Burns and Wholey 1992, pp. 48–49).[3]

While the literature on hospital choice contains many studies with conditional logit models (Lee and Cohen (1985); Garnick et al. (1989); Luft et al. (1990); Burns and Wholey (1992)), older studies which explicitly recognize and/or address the IIA problem in hospital choice are scarce. We describe these next, then conclude with more recent studies that have dealt directly with the IIA problem in hospital choice.

McGuirk and Porell (1984) criticize the early gravity model approach (as seen above, this is equivalent to the conditional logit model which includes distance between patient and chosen hospital as an explanatory variable) for ignoring intervening choice alternatives, i.e. treating facilities as though they were in spatial isolation from one another. They cite work by Morrill and Earickson (1968), Roghmann and Zastowny (1979), and Morrill et al. (1970) as examples from the literature which demonstrate the importance of intervening factors (i.e., other available hospitals nearby) on hospital choice. McQuirk and Porell (1984) posit an 'intervening opportunities' model, which is a 'competing destinations model' variant of the more general spatial interactions model. These spatial interaction models describe the flows of patients, as determined by propulsion variables (demand), attraction variables (supply) and spatial separation factors (distance, intervening opportunities, agglomeration effects, and other constraints) (Fotheringham and O'Kelly 1989).

Berry (1994, pp. 246–247) explains that the DRA property will hold for many discrete-choice specifications in which consumer and product characteristics are interacted, allowing consumer differences to have a systematic impact on their preferences. He gives, as examples, studies using consumer data wherein observed consumer characteristics are interacted with product characteristics. In this spirit, a recent paper on hospital choice includes consumers' perceptions of several unobserved hospital attributes such as reputation (Jung et al. 2011). Using a unique approach, the authors use survey data to extract "importance weights" that consumers place on several unobserved attributes when choosing a hospital, and interact these with hospital characteristics. The coefficients of the interaction terms are then interpreted as the perceived amount of each unobserved attribute offered by each hospital. This effectively ensures that consumer differences in perception of the same hospital's attributes will systematically impact their preferences, which addresses the IIA property.

We note here that the distance between a consumer and their chosen hospital (d_{ij}) is *not* such an interaction; it is a consumer's perception regarding the product (hospital) attribute, which is invariant across all consumers of that type (i.e. at

[3] The test compares estimated parameters and covariance matrices from the full choice set (conditional logit model) with the restricted choice set (nested logit model). The test can fail for reasons besides IIA and can yield a negative test statistic (Burns and Wholey 1992, p. 49, footnote 5).

that location). Conditional logit models for hospital choice that simply include distance between patients at a given location and chosen hospital as an explanatory variable do not reflect interaction between the consumer and product characteristics, thus do not address the IIA problem. The first study in the hospital choice literature to incorporate a true distance-based consumer-product interaction was by Burgess and DeFiore (1994), who interacted consumer age with distance to hospital. This introduces heterogeneity across individuals at location i in their perception of the distance to hospital, and thus qualifies as a DRA-type model. Another more recent hospital choice study has interacted distance with personal characteristics, while never discussing the IIA problem, provides an approach to resolve it (Escarce and Kapur 2009).

A more recent paper by Kessler and McClellan (2000) in the hospital-choice literature used a competing-destinations approach, incorporating accessibility factors into the choice model to address the IIA properties. In this paper, the degree of similarity and dissimilarity among choices (hospitals) was modeled by using relative distances among them faced by consumers. The authors modeled consumer choice of hospital as one component of a broader model of hospital competition. On page 586 of their paper, the authors described how including distances to intervening alternatives addresses the IIA property that arises in these types of models. Their systematic parameterization of the spatial structure among competing alternatives was a novel and robust new foundation for measuring hospital market share and also deriving the Herfindhal-Hirschman Index measure of hospital concentration from these market share measures.[4] Another recent paper by Capps et al. (2010) also focused specifically on the similarity/dissimilarity of hospital service availability based on patient's characteristics. Capps et al. used the service match variable to address the fact that Individuals were less likely to elect to receive care at hospitals that did not offer the services they need. Neither the Kessler and McClellan (2000) or Capps et al. (2010) papers reference the spatial modeling literature described above, thus their contributions to the hospital choice literature is at risk of being ignored, or at best lost in the larger contribution of their papers.

More recently, Lee et al. (2012) use an approach with random error components to solve the product substitution problem noted by NEIO researchers (Berry (1994); Berry et al. (1995); Nevo (2001)). To solve the problem of empirical intractability, they adopt the grouped-logit Dirichlet multinomial regression, a recent refinement of the McFadden (1974) model that includes a group level random effect which may correct for omitted group level factors influencing hospital choice (Guimaraes and Lindrooth 2007). Groups are defined to reflect types of people at certain locations, allowing for different types at each location which are interacted with distance to hospital, in the spirit of Burgess and DeFiore (1994). This paper thus provides a complex measure of accessibility based on random utility theory, in

[4] The robustness stems from the fact that the concentration measure is based only on the observable, exogenous characteristics of patients and hospitals – which is a significant improvement over endogenous measures based on shares of revenues or admissions.

which the probability of a person making a particular hospital choice depends on the utility of that choice relative to the utility of all available choices.

More recent literature has adopted even more sophisticated modeling, including the mixed logit specification (Varkevisser et al. 2012),which was used to address the IIA problems inherent in McFadden's classic conditional logit specification (McFadden 1974). These capabilities have emerged with expansions in computing capabilities and software. The mixed logit model is a highly flexible model that can approximate any random utility model (McFadden and Train 2000; Train 2009). The vector of coefficients representing the patient's tastes for travel time and hospital attributes are allowed to vary with patients according to an assumed random probability distribution, thus it is not necessary to interact hospital attributes with patient characteristics to represent heterogeneity among individuals. The authors argue that empirical evidence suggests that the approach using interacted patient and provider attributes to capture patient-level heterogeneity in preferences may only partially account for this heterogeneity (Hole 2008). They argue that the mixed logit approach is preferred because preference heterogeneity that is unrelated to observed patient characteristics would be captured, whereas it would be omitted in the modeling with direct interactions.

7.6.1 Policy Application of Empirical Estimates

Assuming that individuals assign utilities to all available choices, and select the most attractive choice destination, then potential accessibility can be defined as the denominator of the multinomial logit model, or logsum (Handy and Niemeier 1997). The logsum is a summary measure indicating the desirability of the entire choice set, where the utility function includes attributes of the person, the available destinations, transportation impedance, and socioeconomic or sociocultural conditions in the person's neighborhood. An empirical advantage of the logsum measure can be derived from an empirically valid conditional logit model, structured in such a way as to address the IIA property, which directly reflects the Dependence of Relevant Alternatives (DRA) that characterize consumer choice in spatially-dependent markets. As noted above, Fotheringham (1983) and Fotheringham and O'Kelly (1989) argue that inclusion of accessibility factors such as travel impedance, or the distance distribution/spatial structure among all alternatives in a conditional logit model deals with the IIA property. In the Lee et al. (2012) paper, when a hospital enters, exits, changes services or ownership – the spatial structure among alternatives changes, directly impacting the odds of choice among any two alternatives. Thus the logsum from a competing-destinations variant of the conditional logit hospital choice model is both empirically valid and has a well-grounded theoretical interpretation.

The major theoretical advantage of the logsum measure of potential accessibility is that, because it is grounded in social choice theory, changes in the accessibility measure over time can be interpreted directly as changes in consumer surplus/social welfare (Handy and Niemeier 1997). For example, Town and Liu (2003) used this approach to determine the social surplus attributable to increased choice among Medicare HMO options in certain regions. In recent hospital choice literature, changes in the logsum have been used to evaluate the effects of major structural changes in health care delivery, such as inclusion of hospitals in a health plan network or rural hospital closure (Rosero-Bixby (2004); Handy and Niemeier (1997); McNamara (1999); Town and Liu (2003); Town and Vistnes (2001); Capps et al. (2010)). This practical application of the empirical findings from choice models is expected to be useful in the evaluation of changes in availability of many types of public infrastructure in the future. The older IIA type models did not provide reliable parameter estimates to use in policy assessments of the value of changing choice options. Thus the evolution of modeling to include DRA formulations has been a significant advance for applied policy analysis.

7.7 Summary

In this paper we compare and contrast two different approaches to building in 'dependence of relevant alternatives' (DRA) in consumer choice models. One approach derives from the regional science literature (Fotheringham's competing destinations model), which has a more spatial orientation; the other derives from more traditional economic models of consumer choice (McFadden's conditional logit and the NEIO random coefficient variants thereof). The major distinction between the two model classes is simply that in the former, spatial structure/ accessibility factors distinguish products, while in the latter, consumer-specific perceptions of dissimilarities in product attributes accomplish this. In fact, the NEIO models are very similar in spirit to the earlier product-attribute-variant of the competing destinations models described by Fotheringham (Meyer and Eagle (1982); Borgers and Timmermans (1987)). We discuss recent hospital choice literature that begins to blend the two approaches, including both spatial interaction and random error components in attempts to model the structural dependence of the set of relevant alternatives to consumers. After discussing even more recent developments, we concluded with an explanation of how results from these DRA models can be used to construct measures that provide reliable information to policy makers about the value or loss to consumers from changes in public infrastructure and facilities. These DRA models will be valuable for public policy evaluations and conclusions regarding the social benefit of changes in rapidly transforming sectors of the economy. At present, the healthcare sector is in the midst of substantial change due to the implementation of US health care reform, which will ultimately affect consumer choices in relation to their insurance coverage, the range of health services they seek, and the health providers they select to deliver these services.

References

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *Rand Journal of Economics, 25*, 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica, 63*(4), 841–890.

Borgers, A., & Timmermans, H. (1987). Choice model specification, substitution, and spatial structure effects. *Regional Science and Urban Economics, 17*, 29–47.

Bresnahan, T., Stern, S., & Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *Rand Journal of Economics, 28*, S17–S44.

Burgess, J., & DeFiore, D. (1994). The effect of distance to VA facilities on the choice and level of utilization of VA outpatient services. *Social Science and Medicine, 39*(1), 95–104.

Burns, L., & Wholey, D. (1992). The impact of physician characteristics in conditional choice models for hospital care. *Journal of Health Economics, 11*(1), 43–62.

Capps, C., Dranove, D., & Lindrooth, R. C. (2010). Hospital closure and economic efficiency. *Journal of Health Economics, 29*(1), 87–109.

Escarce, J. J., & Kapur, K. (2009). Do patients bypass rural hospitals? Determinants of inpatient hospital choice in rural California. *Journal of Health Care for the Poor and Underserved, 20* (3), 625–644.

Feldman, R., Finch, M., Dowd, B., & Cassou, S. (1989). The demand for employment-based health insurance plans. *Journal of Human Resources, 24*, 115–142.

Fotheringham, A. S. (1983). A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning A, 15*, 15–36.

Fotheringham, A. S. (1986). Modeling hierarchical destination choice. *Environment and Planning A, 18*, 401–418.

Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial interaction models: Formulations and applications*. Boston: Kluwer.

Garnick, D., Lichtenberg, E., Phibbs, C., Luft, H., Peltzman, D., & McPhee, S. (1989). The sensitivity of conditional choice models for hospital care to estimation technique. *Journal of Health Economics, 8*(4), 377–397.

Guimaraes, P., & Lindrooth, R. (2007). Controlling for over dispersion in grouped conditional logit models: A computation simple application of Dirichlet Multinomial Regression. *Econometrics Journal, 10*, 439–452.

Handy, S., & Niemeier, D. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A, 29*, 1175–1194.

Hausman, J., & McFadden, D. (1984). A specification test for the multinomial logit model. *Econometrica, 52*(5), 1219–1240.

Hole, A. R. (2008). Modeling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics, 27*(4), 1078–1094.

Jones, A. (2000). Health econometrics. In J. P. Newhouse & A. J. Cutler (Eds.), *Handbook of health economics*. Elsevier: North Holland.

Jung, K., Feldman, R., & Scanlon, D. (2011). Where would you go for your next hospitalization? *Journal of Health Economics, 30*, 832–841.

Kessler, D., & McClellan, M. (2000). Is hospital competition socially wasteful? *The Quarterly Journal of Economics, 115*(2), 577–615.

Lee, H., & Cohen, M. (1985). A multinomial logit model for the spatial distribution of hospital utilization. *Journal of Business and Economic Statistics, 3*(2), 159–168.

Lee,W., Bazzoli, G., Hsieh, E. H., & Mobley, L. (2012). *The effect of hospital safety net contractions on the access to care of uninsured and medicaid populations*. Working paper presented at the 8th World Congress on Health Economics in Toronto.

Luft, H., Garnick, D., Mark, D., Peltzman, D., Phibbs, C., Lichtenberg, E., & McPhee, S. (1990). Does quality influence the choice of hospital? *Journal of the American Medical Association, 263*(21), 2899–2906.

Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic.

McFadden, D. (1978). Modeling choice of residential location. In A. Kalgvist et al. (Eds.), *Spatial interaction theory and planning models*. Amsterdam: North-Holland.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski & D. McFadden (Eds.), *Structural analysis of discrete data: With econometric applications*. Cambridge: MIT Press.

McFadden, D., & Reis, F. (1975). *Aggregate travel demand forecasting from disaggregated behavioral models* (Transportation Board Research, Record No. 534). http://eml.berkeley.edu/reprints/mcfadden/aggregate_disaggregate.pdf

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447–470.

McGuirk, M., & Porell, F. (1984). Spatial patterns of hospital utilization: The impact of distance and time. *Inquiry, 21*(1), 84–95.

McNamara, P. E. (1999). Welfare effects of rural hospital closures: A nested logit analysis of the demand for rural hospital services. *American Journal of Agricultural Economics, 81*, 686–696.

Meyer, R., & Eagle, T. (1982). Context-induced parameter instability in a disaggregated-stochastic model of store choice. *Journal of Marketing Research, 19*, 62–71.

Morrill, R., & Earickson, R. (1968). Variation in the character and use of Chicago area hospitals. *Health Services Research, 3*, 224–238.

Morrill, R., Earickson, R., & Rees, P. (1970). Factors influencing distances traveled to hospitals. *Economic Geography, 46*(2), 161–172.

Nevo, A. (2000). A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of Economics and Management Strategy, 9*(4), 513–548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica, 169* (2), 307–342.

Roghmann, K., & Zastowny, T. (1979). Proximity is a factor in the selection of healthcare providers: Emergency room visits compared to obstetric admissions and abortions. *Social Science and Medicine, 13D*(1), 61–69.

Rosero-Bixby, L. (2004). Spatial access to health care in Costa Rica and its equity: A GIS-based study. *Social Science and Medicine, 58*, 1271–1284.

Town, R., & Liu, S. (2003). The welfare impact of medicare HMOs. *RAND Journal, 34*(4), 719–736.

Town, R., & Vistnes, G. (2001). Hospital competition in HMO networks. *Journal of Health Economics, 20*, 733–753.

Train, K. (2009). *Discrete choice models with simulation* (2nd ed.). New York: Cambridge University Press.

Varkevisser, M., Van der Geest, S., & Schut, F. (2012). Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics, 31*, 371–378.

Chapter 8
Bringing Together Spatial Demography and Political Science: Reexamining the Big Sort

David Darmofal and Ryan Strickler

8.1 Introduction

Scholars often express concerns that researchers are increasingly segregating themselves into silos – despite common concerns and interests, true interdisciplinary research is too often a rarity. Whether due to career incentives that promote research within disciplines but not across, discipline-specific nomenclature, or other factors commonly captured in the term "the sociology of science", researchers too often miss opportunities for cross-disciplinary intellectual fertilization. Consider, for example, the issue of migration. Demographers have long been attuned to issues of migration, particularly international migration. Here, macro- and micro-economic conditions, age cycles, and community ties based on race or ethnicity have been found to play key roles in explaining why individuals migrate – and why they don't. Contemporaneous with this burgeoning interest in migration in demography, political scientists have also been increasingly concerned with issues of migration. Here, the focus is on the political determinants or effects of migration – the sorting of individuals into distinct partisan locales, blue ones for Democrats and red ones for Republicans. The parallel, but separate, tracks of migration research in these two disciplines have thus far produced two principal conclusions: while individuals often migrate, ideas rarely do between disciplines. Mid-level theorizing in both disciplines could benefit from considering how demographers and political scientists can increasingly speak to each other over the shared concern of migration.

Bill Bishop's 2008 book, *The Big Sort: Why the Clustering of Like-Minded America is Tearing us Apart* provides fruitful ground for promoting mid-level theorizing in these two disciplines by simultaneously taking seriously in a popular setting concerns that are central to both disciplines, but also leaving the nuts and

D. Darmofal (✉) • R. Strickler
Department of Political Science, University of South Carolina, Columbia, SC, USA
e-mail: darmofal@gmail.com; ryan.strickler@gmail.com

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_8

bolts linkages between these two disciplines unexplored. As a consequence, a reexamination of *The Big Sort*'s arguments and empirical claims that takes seriously the concerns of both demographers and political scientists can itself form a central and much needed bridge between these two disciplines. This chapter provides a bridge for promoting theoretical development and dialogue between these two disciplines by using Bishop's argument as a venue for exploring the shared interests of these disciplines and examining how work in both can be advanced by taking seriously the timely issue of politically- correlated migration in the United States.[1]

Our chapter is structured as follows. We first detail the central claims made by Bishop in *The Big Sort*. Next, we place Bishop's claims in context by examining the political science literature on political polarization and geography. Next we probe further on the role that migration may play in producing a geography-based political polarization. After examining critiques leveled by political scientists against the analysis in *The Big Sort*, we next examine the quite limited consideration of migration studies in Bishop's book. Here, we identify four central limitations in the book that are produced by this inattention to migration studies. We conclude by examining the opportunity that *The Big Sort* and its arguments provide for the movement away from research silos and toward greater interdisciplinary research on migration-induced political polarization.

8.2 The Big Sort

Bishop's book has drawn considerable interest from scholars and pundits alike. Tapping into provocative questions of political polarization and the electoral impact of migration patterns, the book has led no less a figure than former President Bill Clinton to regularly extol its importance in public speeches. At the 2008 Aspen Ideas Festival, for example, Clinton enthusiastically approved of the book's central argument that "we are growing more isolated in our communities because we are living more and more only with people we agree with," concluding that "this is not good in a democracy," ("A Conversation with President Bill Clinton" 2013).

At heart, Bishop (2008) merges concerns of demographers and political scientists, examining how politically-correlated migration is reshaping communities

[1] Lesthaeghe's Second Demographic Transition thesis provides an existing, critically important linkage between these two disciplines. Lesthaeghe (2010, 1–211) argues that in contrast to the First Demographic Transition (FDT) that occurred in Western countries beginning in the eighteenth century, the Second Demographic Transition (SDT) that began in the 1950s brought "sustained sub-replacement fertility, a multitude of living arrangements other than marriage, the disconnection between marriage and procreation, and no stationary population". Lesthaeghe and Neidert (2006, 2009) find a strong relationship between the SDT and the spatial patterns in voting that are the focus of Bishop's work, and particularly find that blue states and counties are more likely to exhibit features of the SDT than are red states and counties.

throughout the United States. The author argues that, in an increasingly mobile, affluent country, “prosperity and opportunity” allow people to order “their lives around their values, their tastes, and their beliefs,” (12). This creates an “unconscious decision to cluster in communities of like mindedness” (15), thus perpetuating a “giant feedback loop” (39) of homogenizing political discourse.[2] To buttress these claims, Bishop presents evidence showing that:

- The number of “tipped” counties, or counties that consistently voted for one party for President for decades, has increased since World War II
- Nearly two thirds of counties have become less competitive in Presidential elections since 1976
- “Strong Democrat” and “strong Republican” counties have markedly different demographics, religiosities, and opinions on the war in Iraq and homosexuality
- Differences between “strong Democrat” and “strong Republican” counties on educational levels, race, religiosity, and immigrant levels have been growing over time.

Bishop argues that “the sort” is driven by two factors. First, beginning with the social and political upheavals of the 1960s, there has been a “silent revolution,” where people began placing less trust in traditional institutions that have long moored American society: governments, traditional religious denominations, and civic organizations (drawing from Putnam 2000). At the same time, this thesis argues, people became less willing to participate in the “elite driven” politics of traditional parties and more inclined to espouse “a politics of self- expression” (Bishop 2008, 85). As a result, parties increasingly adopted social cleavages (most notably with the rise of the “religious right” in the late 70s), and partisanship has increasingly become a reflection of self-expression.

Second, the geographic dimensions of this political sorting have been fueled by economic mobilization. Since the mid-1960s, America has witnessed a “post-materialist Tiebout migration based on non-economic goods, as people have sought out places that best fit their ways of life, their values, and their politics” (199). Young, educated Democrats are pulled towards “high-tech” cities such as Austin, San Francisco, or Portland, while Republicans congregate in small towns or “low-tech” cities such as Birmingham and Cincinnati. Economically, “high-tech” cities developed the social capital necessary to fuel what Richard Florida describes as “spiky” growth based on “creative class” innovations (cited in Bishop 2008, 131). Politically, these localities forged ever sharper distinctions based on culture and politics. Through this sorting “feedback loop,” the political consequences include gridlock in Washington, ideological “democratic experiments” (300) at the local level, and targeted campaigns focused on rallying the base as opposed to swaying the other side. Bishop claims further that the effect goes beyond politics, as

[2] Research in political communication (Jamieson and Cappella 2008; Lee and Cappella 2001), indicates that this feedback loop may be further promoted by exposure to conservative talk shows such as Rush Limbaugh’s.

churches, advertising, and even philanthropy have become balkanized in a "sorted" America.

8.3 Polarization and Geography

Bishop's concern for polarization, the divergence of political elites and/or the public into distinct, ideologically homogeneous factions, is reflective of a growing concern in both popular and academic discourse. Mainstream news sources, as well as punditry of various political stripes, point to gridlock in Washington, election maps marked by "red" and "blue" states, and movements such as the Tea Party and Occupy Wall Street to posit that the United States is becoming increasingly polarized. The result, the media say, is political dysfunction, intra-group homogeneity, disillusionment with politics, and even the erosion of familial and friend relations (Glass 2012).

Political scientists have long been interested in the concept of polarization; scholars paint a more nuanced picture, though, equivocating from the familiar line that America is irrevocably becoming a country divided into two political nations. There is a strong academic consensus that political elites today are increasingly separated into ideologically homogenous, distinct camps (Hetherington 2001; Bartels 2000; Fiorina 2005; Abramowitz and Saunders 2008). Some scholars have argued that this separation of elite politics has diffused into the polarization of public opinion. Abramowitz (2010, 594), for example, analyzes election and National Election Studies (NES) data to portray a "deepening red-blue divide" at both the state and county level. Campbell (2008) echoes the argument of Abramowitz, pointing to NES ideological and partisan self-identification measures to argue that the populace is deeply split, and political discourse will thus continue to be heated for some time to come. In addition to observational studies, both Levendusky (2009) as well as Druckman et al. (2013) provide experimental evidence that elite polarization can create cues that cause segments of the public to move to one ideological extreme or another.

Other scholars, however, have questioned the polarization thesis. Many suggest that current discourse looks polarized only when one examines a narrow time frame, as opposed to the whole of American history (Fischer and Mattson 2009; Ansolabehere et al. 2006). Others have questioned the extent to which elite polarization has diffused to the public. Fiorina (2005, 13), for example, argues that Americans are "closely but not deeply divided." As political elites are separating ideologically, the public is, to a certain degree, shuffling parties without significantly changing ideological dispositions. Even the so-called "culture wars" are more a reflection of candidates, rather than the public, increasingly adopting divergent positions on social issues. Levendusky (2009) comes to a similar conclusion, drawing a distinction between the "polarization" of elites and the "sorting" of the public. As elites have become polarized, he argues, the masses respond to the clearer, sharper elite cues and "align their partisan and ideological beliefs accordingly" (2). Finally, Carsey and Layman (2002, 788) argue that "many, and perhaps most, citizens are unlikely to respond to political cues provided by party elites because they pay little attention to elite-level politics, because they have no ties or only weak ties to a political party, or both."

"Polarization" can take on many forms: polarization between parties, between age groups, between ethnicities, etc. One prominent variant of the debate is whether America is exhibiting geographic polarization, captured in popular imagination by the divide between "red" and "blue" states or counties. Many argue that geographic polarization is more hype than reality. For example, Ansolabehere et al. (2006) as well as Fiorina (2005) point to NES data showing that respondents in "red" states are very similar in ideological self-placement and opinions on issues to respondents in "blue" states (across both economic and social issues). Likewise, Evans (2003) finds that, while ideological and issue attitudes among Democratic and Republican identifiers have diverged, political attitudes across geographic regions have actually converged. Relatedly, Morrill et al. (2007, 549) argue that "while the polarization version of electoral geography is accurate, it is misleading," showing that there is significant nuance to the picture of rural "red counties" and urban "blue counties".

Nivola and Galston (2008, 236), however, point to bitterly contested primaries and the decline of split-ticket voting to suggest that the electorate is "clustering in 'red' and 'blue' counties, if not states or regions." This argument is picked up by Gimpel and Schuknecht (2003, 1), who argue that, since the founding of America, federalism "acts against unity, making a political system a barrier to homogeneity". Through an examination of voting trends across 12 states, the authors show that political opinions, political cultures, and even epistemologies of words like "Democrat" or "conservative" vary greatly both between and, importantly, within states. They also posit that opinion change in a locality is driven by four factors that interact to varying degrees in different locales: conversion of opinion, mobilization of a previously inactive public, generational change, and in or out-migration.

8.4 Migration and Politics

Gimpel and Schuknecht argue that this last factor, migration, "has been the most important force shaping the political identity of regions," (27). Acknowledgement of the political effects of migration has a deep history, drawing the attention of researchers such as V.O. Key and Phillip Converse. However, some scholars have suggested that migration effects are currently too often overlooked in the public opinion literature (Gimpel and Schuknecht 2003; Jurjevich and Plane 2012; Robinson and Noriega 2010). Given the decline in fertility rates in the United States, "population redistribution trends are increasingly dependent on migration" (Johnson et al. 2005, 791). Thus, trends in electoral change may be increasingly dependent on migration; moreover, with the influx of migrants to states like Florida and North Carolina, the potential for migration to redraw the electoral map may continue in the future (Jurjevich and Plane 2012, 429–430).

Scholars studying migration posit two distinct approaches through which migration could contribute to local political opinion change – compositional and contextual approaches. A compositional approach suggests that opinion is determined by specific demographic characteristics in a locality – age, race, income, etc. Thus, the political effect of migration can be determined by tallying changes in a myriad of relevant demographic variables.[3] A contextual approach, on the other hand, accounts for political socialization and “neighborhood effects,” which exert influence beyond the demographic makeup of individual migrants. While not denying the effect of place, Gimpel and Schuknecht focus primarily on compositional effects of political migration, as they are directly observable and do not rely on vague or untestable notions of “context.” Using this compositional approach, they argue that most political variation between localities can be explained by (1) ideology and issue salience, (2) economic stratification, (3) ethnicity and religion, and (4) race. Likewise, Jurjevich and Plane adopt the compositional approach, critiquing past electoral studies researchers for their (a) inability to disaggregate migration from broader demographic change, (b) inattention to migrant origins as well as destinations, and (c) assumption that migrants are predominantly Republican.[4] Using US Census data from 1995 to 2000, they show that migration leads to “an increased, but varied ‘potential purpleness’ of the electorate” at the state level (442), with streams of migrants contributing to both the strengthening and diluting of the parties’ strength across states in complex ways.

While acknowledging the value of the compositional approach in elucidating the intersection of migration and opinion, Brown (1988) argues that the “contextual” approach is too often ignored in the literature. Critical of past literature that ignores the effect of migration on the migrant, as well as assumes that migrants have a degree of “psychological immunity” to countervailing political messages (14), he argues that “few (migrants) ever have the resources to remain steadfast on their partisan and political beliefs when everything around them has changed” (15). Comparing the effect of “political environment” to migrants’ voting behavior and opinion, he shows that a migrant’s current, not previous, political environment is the primary driver of voting decisions and partisanship. Likewise, Huckfeldt et al. (1995) argue that the effect of a migrant’s political environment is mediated through the “weak” social ties he or she develops. As an individual interacts with others outside his or her immediate social cohort, the authors empirically demonstrate that his or her political opinion will more closely match that of the larger community. Furthermore, McKee and Teigen (2009) ascribe importance to the contextual effect of “place,” viewing it as a conduit through which specific, measurable location characteristics impact opinion (485). Using 2000 and 2004

[3] Of course, it is impossible to determine and measure every demographic characteristic that contributes to political opinion. The compositional approach only suggests that these sort of variables, if they all could be measured, could perfectly explain change in public opinion (without relying on “socialization” or “contextual” effects).

[4] Examples of this argument date back to Campbell et al’s *The American Voter* (1960).

Presidential election data, they show that population density (measured as urban, rural, or suburban) and region both independently influenced voting behavior; the effect of population density varied by region, and the effect of region varied by level of population density.

8.5 Struggling with *The Big Sort*

If Bishop (and Bill Clinton) are correct, the effects of internal migration (be they compositional or contextual) are creating a "post-materialist" polarization. The consequences are dire:

> balkanized communities whose inhabitants find other Americans to be culturally incomprehensible; a growing intolerance for political differences that has made national consensus impossible; and politics so polarized that Congress is stymied and elections are no longer just contests over policies, but bitter choices between ways of life, (Bishop 2008, 14).

To be fair, some scholars see some potential positives in this sort of "sorting;" Levendusky (2009), for example, argues that partisan "sorting" helps voters "participate more effectively" as democratic citizens by giving them clear, meaningful choices at the ballot box (140). Whether positive or negative, though, the significant impact of an alleged "Big Sort" necessitates careful scrutiny of the argument provided by Bishop.

Unfortunately, a number of methodological and conceptual issues can be raised, drawing from both political science and demography literatures. From the political science literature, first, a number of scholars have taken issue with the time frame Bishop uses, suggesting that Bishop's focus on the post-WWII era, and particularly 1976–2004, paints a misleading picture, as the mid-twentieth century was a unique time of party heterogeneity and relative political détente (Ansolabehere et al. 2006; Abrams and Fiorina 2012; Glaeser and Ward 2006). Second, as discussed previously, many scholars argue that divergent voting behavior does not necessarily indicate real ideological differences in public opinion. For example, Abrams and Fiorina (2012) argue that looking at Presidential election data (as Bishop does), in the context of political *elites* polarizing, skews the perception of *public* polarization upward. Instead, these authors look at county level *voter registration* data in 21 states that record partisan affiliation with registration (a more stable measure over time, they argue). These data show that the number of independents has increased dramatically since 1976, suggesting that the public is not echoing polarization at the elite level. McGhee and Krimm (2009) likewise analyze county-level registration data and come to a similar conclusion.

Third, Bishop focuses almost exclusively on culture and lifestyle as factors driving polarization in contemporary America. He is not alone in arguing the increased salience of "post-materialist" social issues; popular books, such as Thomas Frank's *What the Matter with Kansas?* and David Brooks's *Bobos in*

Paradise, argue that rifts in culture, religiosity, and lifestyle – not economic issues – drive liberal and conservative opinion apart in the twenty-first century.

Abramowitz and Saunders (2008), moreover, point to NES data to make the claim that “the religious divide is now much deeper than the class divide” (although he limits his analysis to white voters) (550). For many (perhaps most) other academics, however, the consensus is that economic concerns still hold sway over public opinion. Ansolabehere et al. (2006), as well as Gelman et al. (2008), point to evidence suggesting that economic issues are still top of mind for most Americans, with social issues only having a secondary effect on opinion. Fiorina (2005) likewise argues that religious and social cleavages in society have become salient in addition to, not at the expense of, economic cleavages. Social issues have furthermore only come into salience due to candidates adopting increasingly opposed stances, not due to an increased divergence in public opinion.

8.6 Migration Critiques

Bishop addresses some of the potential critiques leveled by political science scholars in his work; for example, he acknowledges that it is “certainly the case” that sorting would look less pronounced if one took a longer view of American history (25), and he draws form the work of Abramowitz and others to suggest that Fiorina is mistaken in his claim that the United States is “closely, not deeply” divided (25-8). However, even though his thesis *hinges on migration*, he fails to engage migration scholarship in a serious way. Instead, he presents county and metropolitan level data over time, assuming that “post-materialist” migration is fueling the political and cultural sorting he observes. This results in four central limitations of the book: (1) unquestioned assumptions regarding the drivers of migration, (2) inattention to the mechanism that fuels political change, (3) a focus on internal migration to the exclusion of international migration, and (4) an inappropriate level of analysis for studying migration.

Causes of Migration Bishop’s focus on culture and lifestyle as a driver of *migration* (not just geographic polarization, as Abramowitz and Saunders (2008) would assert) is particularly questionable.

He contends that, by the 1990s, “there was a surge of people who wanted to live in cities for what could only be social – or even aesthetic – reasons,” (152). As a result, fostering a particular “lifestyle” has become the city’s modus operandi and key to economic development, in order to lure a fair share of the nomadic, wealth-producing “creative class” (Florida 2002).[5] As evidence, Bishop cites growing

[5] In addition to being a spurious driver of migration, “post-materialist” lifestyle positioning has also been called into question as a driver of local economic development. For a particularly strong critique of the “creative class” thesis, see Peck (2005).

differences in “high tech” versus “low-tech” metropolitan areas with regard to race, age, income, occupation, patent creation, and the “social capital” indicators developed by Putnam (2000).

What Bishop does not sufficiently allow for, however, is the possibility that these demographic indicators, or other indicators suggested by migration scholars, are the potential primary drivers in migration patterns. While recent research in US internal migration patterns has been somewhat sparse,[6] there are studies that suggest that this could be the case. Johnson et al. (2005), for example, examine migration patterns by age cohorts and conclude that there is “a striking consistency in the overall migration signatures of particular types of counties” based on age and life cycle (808). Lee et al. (1994) also find that age, in addition to other individual factors such as tenure in neighborhood and homeowning status, play a much stronger role in predicting migration out of a neighborhood than characteristics (real or perceived) of the neighborhood itself. Other factors, such as migration distance, unemployment, or other economic concerns are not central to Bishop’s analysis, yet may play important roles in shaping the political migration that he documents. Greenwood (1988), for example, cites the growth of the labor force as well as the increasing concentration of employment opportunities in the South and West regions to argue that the population growth of these regions during the 1970s and 1980s was fueled, in part, by domestic migrants seeking work. As another example, Pandit (1997) analyses data from 1949 to 1993 to show that economic conditions interact with the sizes of age cohorts to determine a period’s overall migration rate. (This focus on age effects on migration mirrors Parker’s (2014) insightful analysis of age-specific rates of out-migration among the Karen in Thailand in this book). The effect of both “cohort size” and economic conditions is stronger for long-distance (interstate), as opposed to shorter (intrastate), migrations.

Furthermore, race and ethnicity can also play a critical role in migration. The “Great Migration” of African Americans out of the rural South in the first half of the twentieth century, in part to escape racial prejudice, is well documented (Price-Spratlen 2008; Tolnay et al. 2002). As another example of the intersection of ethnicity and migration, South et al. (2005) show that the propensity of Latinos to move to neighborhoods with a large percentage Anglo population depends upon the migrant’s human and financial capital as well as his/her English language proficiency with important variations in these broad trends for Mexican, Puerto Rican, and Cuban subgroups. In sum, important individual-level correlates of migration are deemphasized in Bishop’s analysis, but are likely to have an important impact

[6] Despite the research discussed here, demographers have perhaps not examined patterns and effects of internal migration in the United States as fully as they could. Ellis (2012), for example, laments the fact that migration scholars have focused on international migration into the US in lieu of internal migration, and discusses ways migration scholars can both transfer international-level analytical tools to internal migration studies as well as link internal and international migration together. Skeldon (2006, 17) also recognizes this shift, arguing that, in migration research, “the word ‘migration’ has come to mean ‘international migration’...”.

on political migration. By looking only at correlations between demographic patterns of "polarized" counties or "high tech" cities, however, one cannot determine a causal relationship.

Effects of Migration If migration is indeed driven by the desire to live in culturally "like" communities, the mechanism of geographic political change is still left unspecified. As discussed earlier, a key distinction in the migration literature is between "compositional" factors of migration-induced change, or political change resulting purely from demographic change, and "contextual" factors, such as the effect of "place" or "political environment" on the new migrants' attitudes (Gimpel and Schuknecht 2003; Brown 1988). Determining when and where "compositional" or "contextual" effects take precedence with political migration is key to making predictions of changes in political geography. As migration into the South and West is predicted to continue into the future (and, by 2030, Florida, California, and Texas account for nearly one half of the US population) (U. S. Census Bureau 2013), will migrants take their politics with them, or will their new environment influence their opinions? The answer to this question is key to anticipating changes in the electoral landscape; however, the observational data presented by Bishop is silent in this regard.

Connection Between Immigration and Internal Migration The domestic migration that Bishop focuses on does not happen in a vacuum, as international migration has a profound effect on internal population flows. For example, African American migration to the North in the twentieth century was partly a response to slowing European immigration, and the flight of agricultural workers to Northern factories during World War II was the impetus for the "Bracero" program, which brought in 4 million Mexican immigrants to work on southwestern farms (Ellis 2012, 198). Citing these historical examples, Ellis argues that a fuller account of internal migration would incorporate "linked-flow studies" of domestic and international migration interactions (197).

The "linked flow" between immigration and internal migration, discussed by Ellis and others (Baines 1985; Ley and Tutchener 2001; Card 2001) has clear political import, as immigration drives political changes along both "compositional" and "contextual" lines. While new immigrants are often restricted from political participation through either legal means (Logan et al. 2009) or discouraged through a lack of political socialization or English proficiency (Cho 1999), they nonetheless harbor views that contribute to the political zeitgeist of a community. Also, with certain electoral conditions and issue cleavages in place, foreign-born immigrants can at times have a significant impact on elections (Barreto 2005; Pantoja et al. 2001). Moreover, second and third generations of immigrants can have tremendous effects on the political composition of a community – the continuing influence of the Cuban community in Miami and Florida state politics is just one indicator of this (Moreno 1997). Compositional effects may be felt away from "immigration gateways" as well, as areas not host to new immigrants, Frey argues, "are becoming more conservative and more likely to vote Republican," (1999, 97).

This, he argues, is driving a "demographic balkanization" (78). It is important to note, however, that in contrast to Frey, Lichter and Johnson (2006, 109) find little evidence of balkanization and conclude that "immigrants are less concentrated today than in the past and they are less segregated from other population groups, including their own racial group and whites."

As a "contextual" effect, there is a deep historical precedent for international immigration provoking political antipathy and anger in "traditional" native-born communities, dating back to the nineteenth century Know-Nothing party. Recently, Parker and Barreto argue that the rise of the Tea Party, part and parcel with this "Know-Nothing" tradition, is driven by a perceived loss of power to a political "Other." This "Other" includes, among other groups, foreign immigrants. In fact, their polling shows that a majority of Tea Party sympathizers feel that immigrants (regardless of legal status) are "too powerful" and "increase crime in America" (2013, 171). As "red" states such as Arizona, Texas, and the Deep South states have received a recent influx of immigrants (Donato et al. 2008), many have channeled this nativist sentiment into passing stringent anti-immigrant legislation (Parker and Barreto 2013, 165; Sabia 2010). As immigrants increasingly move to "non-traditional" destinations (Massey and Capoferro 2008; Hall 2013), reactionary politics in some of these destinations may continue into the future.

International migration also plays a role in Lesthaeghe and Neidert's (2006) analysis of the second demographic transition in the United States, and particularly the seemingly anomalous patterns of demographic transition in the United States vs. many other Western industrial countries. In contrast to the first demographic transition, in which declines in fertility and mortality marked many Western countries beginning in the eighteenth century, the second demographic transition that begin in the 1950s and has spread to many Western industrial countries has been marked by, among other characteristics, a focus on post-materialist concerns and higher-order needs including self-actualization (Inglehart 1970; Maslow 1954, both cited in Lesthaeghe and Neidert (2006, 669)), sub-replacement fertility, and non-marital cohabitation (Lesthaeghe and Neidert 2006, 669; Lesthaeghe 2010, 211).

Viewed in the context of other Western nations' trends, the United States' demographic transition has seemed like an outlier, marked as it is by a fertility rate that actually increased between 1981 and 2001, placing it just above replacement level (Lesthaeghe and Neidert 2006, 670). Lesthaeghe and Neidert (2006, 693–694) trace this higher comparative fertility rate in the United States to the particularly high fertility rate among Hispanics in the country. International immigration of Hispanics who are just completing their first demographic transition has produced a total fertility rate that masks the second demographic transition that has occurred in many areas in the United States (Lesthaeghe and Neidert 2006, 694). They demonstrate that blue states and counties are marked by their sharing many of the attributes of the second demographic transition while red states and counties are marked by a stronger support for the religious right and a lesser reflection of the second demographic transition (Lesthaeghe and Neidert 2006, 684–693).

In sum, international migration has substantial, complex effects on both patterns of political "sorting" and "linked" internal migration flows. By failing to take the international context into account, Bishop misses an important layer of the story that would only strengthen his argument.

Levels of Analysis Problem A key limitation of Bishop's analysis is the use of *aggregate* county and metropolitan statistical area (MSA) analyses to infer individual migration decisions.[7]

First, these geographic units are too large; it is well-known that counties are not viewed by most citizens as their principal social, political, or economic communities. As such, this level of analysis lacks face validity; in making a migration decision, individuals may be motivated to move to Austin or Raleigh, but they would not be motivated to settle in Travis County or Wake County. Bishop's analysis thus operates at a theoretically inappropriate level of aggregation, which could bias his results. Second, because individual-level factors are likely to play a critical role in migration decisions, a causal analysis of migration patterns at any level must incorporate individual level data, or else it runs the risk of erroneously imputing individual-level motivations on migrants.

Recognizing the need to use individual level data to study migration patterns, McDonald (2011) uses 2006 Cooperative Congressional Election Study data, change-of-address data from the US Postal Service, and Presidential election data to show that conservative individuals tend to migrate to "Republican" districts, and liberals to "Democratic" districts. In his paper, he correctly argues that:

> The granularity of locations (studied) will affect any prediction of sorting or convergence. When we examine relatively small area units, such as neighborhoods, or even the suburban component of a metropolitan area, we may find sorting that is undetectable within and between large regions, counties, or states. Our ability to observe and evaluate either sorting or convergence depends completely on the unit of analysis, and the particular consequences also depend completely on how we choose to aggregate. (517).

Despite this statement, though, his dependent variable for migration destination is measured at the Congressional District level, which is often larger than the county-level data used by Bishop and more prone to ecological fallacy. Thus, while he offers an improvement by examining individual-level migrants, his reliance on district-level data weakens his results.

A methodological "gold standard" for research into political migration is offered by Cho et al. (2013). These authors argue that though "county-level results might look suggestive, their relationship with individual-level tendencies might not be in the same direction or of comparable magnitude" (857). As a consequence, these

---

[7] If Bishop were to limit his inferences to those at the aggregate level, his analysis would suffer from the modifiable areal unit problem (MAUP), the fact that aggregate-level findings depend upon the aggregate-level areal units used for analysis. Even limiting one's interest to the aggregate level, there is little reason to believe that counties as arbitrary units drawn for purposes of governmental administration are the appropriate areal units for a study of citizens' chosen local contexts. For discussions of MAUP see Openshaw and Taylor (1979, 1981).

authors examine individual migrant voter registration records to show that a ZIP code's political makeup (measured by differences in Republican and Democratic registration rates) plays a modest role in migration decisions, with secondary factors that are related to individual-level partisanship such as income, race, and population density playing a larger role. Political makeup plays a stronger role in migration for Republicans, as well as those moving longer distances. Cho et al. (2013, 867) argue that these processes, albeit gradual, have the potential to "not only change the political landscape but also create new environments for the socialization of citizens."

In many ways Cho, Gimpel, and Hui's analysis does present a gold standard in the current literature because it links both individual- and aggregate-level data, recognizing that a focus solely on the latter will run into the ecological fallacy if we are interested in explaining individual-level migration decisions while a focus solely on the former will run into the atomistic fallacy if we ignore the effects of factors above the individual level that shape migration patterns. Because of the critical roles that interactions between individual-level and aggregate-level factors are likely to play in influencing migration, multi-level modeling presents a fruitful modeling approach for examining cross-level effects shaping migration patterns.[8]

8.7 Incorporating Space into *The Big Sort*

Political geography is central to Bishop's *The Big Sort* argument and yet, paradoxically, spatial concerns play only a minor role in his account. Although, as discussed above, his county level of analysis has drawbacks, it also is helpful in moving scholars away from the blunt "red state-blue state" dichotomy that ignores substate variation in partisan voting. But in employing a county level of analysis, Bishop too often treats these counties as atomistic entities, ignoring the important question of substate partisan regions (see, e.g., Nardulli 1995).

A critical question regarding The Big Sort is the spatial dimension of this sort. What is the spatial structure of this partisan sorting? Are adjacent counties exhibiting similar patterns of polarization toward the Democratic or Republican Party? Do substate regions of adjacent counties serve as regional magnets for the in- (or out-) migration of Democrats and Republicans? Are patterns of migration marked by spatial dependence and if so, what is the source of this spatial dependence?

All of these questions are important for developing middle-range theory in spatial demography. The sources of any spatial dependence in partisan migration

[8] It is important to incorporate both origin and destination characteristics when modeling migration decisions. If only the latter are modeled, a common flaw in the existing literature, we will be limited in our understanding of how individuals drawn from particular origin locales are drawn to particular destination locales. See Pelligrini and Fotheringham (1999) for an important discussion of this concern (see also Farmer 2011).

and polarization are particularly consequential. Regions comprised of adjacent counties with similar patterns of partisan polarization may exhibit this spatial dependence for either of two principal reasons. On the one hand, citizens in neighboring counties may exhibit similar movement toward the Democratic or Republican Party due to a process of behavioral diffusion, in which political conversations promote political polarization. If such behavioral diffusion occurs across neighboring counties, this would produce a spatial lag process that should be modeled via a spatially lagged dependent variable. Alternatively, it may be that the neighboring counties exhibit little interaction, but instead are serving as magnets for the in- or out-migration of Democrats or Republicans due to exposure to common external shocks, such as the decay of old industries or the rise of a knowledge economy. If so, these shared external shocks would be modeled via a spatial error model. Determining which of these processes is producing spatial dependence in partisan polarization among neighboring counties is critically important for understanding how migration is spurring partisan polarization in the United States.

8.8 Conclusion

Bill Bishop's book, *The Big Sort*, is an interesting, provocative book, a compelling tale of warring "red" and "blue" communities, driven apart by the decline of traditional institutions, "creative class" migration patterns, and the growth of a culture and politics of "self-expression." This chapter is not necessarily arguing that Bishop is incorrect; America could possibly be fragmenting, and this fragmentation could, as Bishop (2008, 199) argues, possibly be driven by a "post-materialist Tiebout migration based on non-economic goods." However, the evidence provided by Bishop – aggregate, county or MSA-level trends in demographics, partisanship, and public opinion – is not enough to give us a definitive answer. It opens the door for political scientists, such as Fiorina and Abrams (2012), to challenge his thesis that America is increasingly polarized.

Further, it opens the door for migration scholars to offer other causal factors for "sorting" established in the literature, including age, economic factors, and race or ethnicity. More broadly, to draw a valid causal inference concerning the political effect of migration, one cannot employ broad, aggregate data. One must match individual, migrant-level data with data on both the origin and destination political environments below the county level. Cho, Gimpel, and Hui's work (2013), which combines individual level migrant data with ZIP code political environment data, is an example in this regard.

The Big Sort gives us the opportunity to reflect on the "silos" academics often find themselves in, as well as opportunities to benefit from interdisciplinary inquiry. The phenomenon of migration is a perfect example of this. Gimpel and Schucknect (2003, 27) argue that although migration has long been overlooked in political science studies of local politics, it plays a critical role in shaping these politics.

Bishop's work likewise sees migration driving local politics; indeed, for him, it is creating a dangerous level of geographic polarization. Greater engagement with the work of migration scholars could allow him to make stronger, and more nuanced, claims regarding the causes and process mechanisms (compositional or contextual), driving migration-fueled polarization. Given the dearth of recent research in internal migration (see Ellis 2012), one intriguing possibility would be to look into international migration research, seeing what concepts, methodological tools, and interconnections can be appropriated for the study of domestic movement and political change.

In the end, all those that study polarization, be they demographers, political scientists, or those from other fields, could benefit from approaching polarization from an interdisciplinary perspective. If it is true, as Bishop states in his title, that "the clustering of like-minded Americans is tearing us apart," the need for this is dire. Ironically, the clustering of like-minded scholars – demographers and political scientists engaging in a closer, more fruitful dialogue – may provide us with insights that can help remedy the negative effects of geographic polarization.

References

Abramowitz, A. I. (2010). Transformation and polarization: The 2008 Presidential election and the new American electorate. *Electoral Studies, 29*(4), 594–603.

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *Journal of Politics, 70*(2), 542–555.

Abrams, S. J., & Fiorina, M. P. (2012). 'The big sort' that wasn't: A skeptical reexamination. *PS: Political Science and Politics, 45*(2), 203–210.

Ansolabehere, S., Rodden, J., & Snyder, J. M. (2006). Purple America. *The Journal of Economic Perspectives, 20*(2), 97–118.

Aspen Ideas Festival. (2013). *A conversation with President Bill Clinton*. http://www.aspenideas.org/session/conversation-president-bill-clinton. Accessed 13 Mar 2013.

Baines, D. (1985). *Migration in a mature economy: Emigration and internal migration in England Wales, 1861–1900*. Cambridge: Harvard University Press.

Barreto, M. A. (2005). Latino immigrants at the polls: Foreign-born voter turnout in the 2002 election. *Political Research Quarterly, 58*(1), 79–86.

Bartels, L. M. (2000). Partisanship and voting behavior, 1952–1996. *American Journal of Political Science, 44*, 35–50.

Bishop, B., with Cushing, R. G. (2008). *The big sort: Why the clustering of like-minded Americans is tearing us apart*. Boston: Mariner Books.

Brown, T. A. (1988). *Migration and politics: The impact of population mobility on American voting behavior*. Chapel Hill: The University of North Carolina Press.

Campbell, J. E. (2008). Polarization runs deep, even by yesterday's standards. In P. S. Nivola & D. W. Brady (Eds.), *Red and blue nation? Characteristics and causes of America's polarized politics, volume one* (pp. 152–162). Washington, DC: Brookings Institution Press.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics, 19*, 22–64.

Carsey, T. M., & Layman, G. C. (2002). Party polarization and 'conflict extension' in the American electorate. *American Journal of Political Science, 46*(4), 786–802.

Cho, W. K. T. (1999). Naturalization, socialization, participation: Immigrants and (non-) voting. *Journal of Politics, 61*(4), 1140–1155.
Cho, W. K. T., Gimpel, J. G., & Hui, I. S. (2013). Voter migration and the geographic sorting of the American electorate. *Annals of the Association of American Geographers, 103*(4), 856–870.
Donato, K. M., Tolbert, C., Nucci, A., & Kawano, Y. (2008). Changing faces, changing places: The emergence of new nonmetropolitan immigrant gateways. In D. S. Massey (Ed.), *New faces in new places: The changing geography of American immigration* (pp. 75–98). New York: Russell Sage.
Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review, 107*(01), 57–79.
Ellis, M. (2012). Reinventing US internal migration studies in the age of international migration. *Population, Space and Place, 18*(2), 196–208.
Evans, J. (2003). Have Americans' attitudes become more polarized? – An update. *Social Science Quarterly, 84*(1), 71–90.
Farmer, C. (2011). *Commuting flows & local labour markets: Spatial interaction modelling of travel-to-work*. Ph.D. dissertation, National University of Ireland.
Fiorina, M. P. (2005). *Culture war? The myth of a polarized America*. New York: Pearson Education.
Fischer, C. S., & Mattson, G. (2009). Is America fragmenting? *Annual Review of Sociology, 35*(1), 435–455.
Florida, R. (2002). *The rise of the creative class*. New York: Basic Books.
Frey, W. H. (1999). Immigration and demographic balkanization: Toward one America or two? In J. W. Hughes & J. J. Seneca (Eds.), *America's demographic tapestry: Baseline for the new millennium* (pp. 78–97). New Brunswick: Rutgers University Press.
Gelman, A., Park, D., Shor, B., Bafumi, J., & Cortina, J. (2008). *Red state, blue state, rich state, poor state*. Princeton: Princeton University Press.
Gimpel, J. G., & Schuknecht, J. E. (2003). *Patchwork nation: Sectionalism and political change in American politics*. Ann Arbor: University of Michigan Press.
Glaeser, E. L., & Ward, B. A. (2006). Myths and realities of American political geography. *Journal of Economic Perspectives, 20*(2), 119–144.
Glass, I. (2012). Red state blue state. *This American life*. Chicago: Chicago Public Media.
Greenwood, M. J. (1988). Changing patterns of migration and regional economic growth in the U. S.: A demographic perspective. *Growth and Change, 19*, 68–87.
Hall, M. (2013). Residential integration on the new frontier: Immigrant segregation in established and new destinations. *Demography, 50*(5), 1873–1896.
Hetherington, M. J. (2001). Resurgent mass partisanship: The role of elite polarization. *American Political Science Review, 95*(3), 619–631.
Huckfeldt, R., Beck, P. A., Dalton, R. J., & Levine, J. (1995). Political environments, cohesive social groups, and the communication of public opinion. *American Journal of Political Science, 39*(4), 1025–1054.
Inglehart, R. (1970). *The silent revolution*. Princeton: Princeton University Press.
Jamieson, K. H., & Cappella, J. N. (2008). *Echo chamber: Rush Limbaugh and the conservative media establishment*. New York: Oxford University Press.
Johnson, K. M., Voss, P. R., Hammer, R. B., Fuguitt, G. V., & McNiven, S. (2005). Temporal and spatial variation in age-specific net migration in the United States. *Demography, 42*(4), 791–812.
Jurjevich, J. R., & Plane, D. A. (2012). Voters on the move: The political effectiveness of migration and its effects on state partisan composition. *Political Geography, 31*, 429–443.
Lee, G., & Cappella, J. N. (2001). The effects of political talk radio on political attitude formation: Exposure versus knowledge. *Political Communication, 18*(4), 369–394.
Lee, B. A., Oropesa, R. S., & Kanan, J. W. (1994). Neighborhood context and residential mobility. *Demography, 31*(2), 249–270.

Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and Development Review, 36*(2), 211–251.

Lesthaeghe, R. J., & Neidert, L. (2006). The second demographic transition in the United States: Exception or textbook example? *Population and Economic Development Review, 32*(4), 669–698.

Lesthaeghe, R., & Neidert, L. (2009). US presidential elections and the spatial pattern of the American second demographic transition. *Population and Development Review, 35*(2), 391–400.

Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. Chicago: University of Chicago Press.

Ley, D., & Tutchener, J. (2001). Immigration, globalisation, and house price movements in Canada's gateway cities. *Housing Studies, 16*, 199–223.

Lichter, D. T., & Johnson, K. M. (2006). Emerging rural settlement patterns and the geographic redistribution of America's new immigrants. *Rural Sociology, 71*(1), 109–131.

Logan, J. R., Oh, S., & Darrah, J. (2009). The political impact of the new Hispanic second generation. *Journal of Ethnic and Migration Studies, 35*(7), 1201–1223.

Maslow, A. H. (1954). *Motivation and personality*. New York: Harper and Row.

Massey, D. S., & Capoferro, C. (2008). The geographic diversification of American immigration. In D. S. Massey (Ed.), *New faces in new places: The changing geography of American immigration* (pp. 25–50). New York: Russell Sage.

McDonald, I. (2011). Migration and sorting in the American electorate: Evidence from the 2006 cooperative congressional election study. *American Politics Research, 39*(3), 512–533.

McGhee, E., & Krimm, D. (2009). Party registration and the geography of party polarization. *Polity, 41*(3), 345–367.

McKee, S. C., & Teigen, J. M. (2009). Probing the reds and blues: Sectionalism and voter location in the 2000 and 2004 US presidential elections. *Political Geography, 28*(8), 484–495.

Moreno, D. (1997). The Cuban model: Political empowerment in Miami. In F. C. Garcia (Ed.), *Pursuing power: Latinos and the political system* (pp. 208–226). Notre Dame: University of Notre Dame Press.

Morrill, R., Knopp, L., & Brown, M. (2007). Anomalies in red and blue: Exceptionalism in American electoral geography. *Political Geography, 26*(5), 525–553.

Nardulli, P. F. (1995). The concept of a critical realignment, electoral behavior, and political change. *American Political Science Review, 89*(1), 10–22.

Nivola, P. S., & Galston, W. A. (2008). Toward depolarization. In P. S. Nivola & D. W. Brady (Eds.), *Red and blue nation? Characteristics and causes of America's polarized politics, volume two* (pp. 235–284). Washington, DC: Brookings Institution Press.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127–144). London: Pion.

Openshaw, S., & Taylor, P. J. (1981). The modifiable areal unit problem. In N. Wrigley & R. J. Bennett (Eds.), *Quantitative geography: A British view* (pp. 60–70). London: Routledge.

Pandit, K. (1997). Cohort and period effects in US migration: How demographic and economic cycles influence the migration schedule. *Annals of the Association of American Geographers, 87*(3), 439–450.

Pantoja, A. D., Ramirez, R., & Segura, G. M. (2001). Citizens by choice, voters by necessity: Patterns in political mobilization by naturalized Latinos. *Political Research Quarterly, 54*(4), 729–750.

Parker, D. M. (2014). Human migration and spatial synchrony: Spatial patterns in temporal trends. In F. M. Howell, J. R. Porter, & S. A. Matthews (Eds.), *Recapturing space*. Berlin: Springer.

Parker, C. S., & Barreto, M. A. (2013). *Change they can't believe in: The tea party and reactionary politics in America*. Princeton: Princeton University Press.

Peck, J. (2005). Struggling with the creative class. *International Journal of Urban and Regional Research, 29*(4), 740–770.

Pelligrini, P. A., & Fotheringham, A. S. (1999). Intermetropolitan migration and hierarchical destination choice: A disaggregate analysis from the U.S. public use microdata samples. *Environment and Planning A, 31*, 1093–1118.

Price-Spratlen, T. (2008). Urban destination selection among African Americans during the 1950s great migration. *Social Science History, 32*(3), 437–469.

Putnam, R. (2000). *Bowling alone: The collapse and revival of the American community*. New York: Simon and Schuster Paperbacks.

Robinson, T., & Noriega, S. (2010). Voter migration as a source of electoral change in the rocky mountain west. *Political Geography, 29*(1), 28–39.

Sabia, D. (2010). The anti-immigrant fervor in Georgia: Return of the nativist or just politics as usual? *Politics & Policy, 38*(1), 53–80.

Skeldon, R. (2006). Interlinkages between internal and international migration and development in the Asian region. *Population, Space and Place, 12*(1), 15–30.

South, S. J., Crowder, K., & Chavez, E. (2005). Migration and spatial assimilation among US Latinos: Classical versus segmented trajectories. *Demography, 42*(3), 497–521.

Tolnay, S. E., Adelman, R. M., & Crowder, K. D. (2002). Race, regional origin, and residence in northern cities at the beginning of the great migration. *American Sociological Review, 67*, 456–475.

U. S. Census Bureau, Demographic Internet Staff. (2013). *2005 interim state population projections*. http://www.census.gov/population/projections/data/state/projectionsagesex.html. Accessed 12 Apr 2013.

Part III
Middle Range Theory in Application

Chapter 9
Demography and Democracy: Exploring the Linkage Between Age and Voter Turnout in Italy with Geospatial Analysis

Michael Shin and John Agnew

9.1 Introduction

Electoral outcomes in both established and new democracies alike generate a flurry of interest and analysis. One of the statistics that frequently accompanies election results is that for electoral participation or voter turnout. Since the right to vote is fundamental to democracy, the extent to which this right is exercised is often used as a barometer of the quality of a democracy or the legitimacy of an election. Despite the spread of democracy around the world, increased opportunities to vote in supra-national, national and sub-national elections, and local efforts to 'get out the vote', since 1990 turnout rates around the world have declined (Fruncillo 2004). Though several explanations for declining rates of voter participation have been put forth, such as voter apathy, weakening levels of partisanship and party system change, the recent drop in turnout in many democracies remains a mystery. Though in US politics older voters are often seen as much more active and likely to vote in a given election than younger ones, elsewhere in the world this relationship does not seem to hold up, at least not to the same extent. Thus, age would seem to be a useful demographic variable to examine in relation to electoral turnout. As populations age, as is characteristic of many countries today, turnout in elections seems to be declining. Is this truly the case?

The reasons to vote or not are varied. Herein lays a large part of the problem in trying to provide a single explanation for declining turnout for an entire country. In different places within a country different mixes of reasons may be at work. Much

M. Shin (✉)
Department of Geography, University of California-Los Angeles (UCLA), 1255 Bunche Hall, Los Angeles, CA 90095-1524, USA
e-mail: shinm@geog.ucla.edu

J. Agnew
Department of Geography, University of California-Los Angeles, Los Angeles, CA, USA

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_9

of the political science literature, particularly in the United States, tends to ascribe non-voting to voting's self-evident "irrationality" for an individual voter engaged in weighing its personal costs and benefits: the fact that a single vote makes very little difference to the final outcome and is thus not worth the effort needed to cast it (e.g. Downs 1957; Riker and Ordeshook 1968). Seeming mass apathy can thus be put in a positive light. Some commentators, however, dispute this, suggesting that non-voting results much more from either increased barriers to voting (as in the historic and contemporary voter suppression efforts by both parties in the United States at different times to bar their opponent's presumed voters from actually voting), the rising number of so-called "ineligibles" in a given potential voting population (non-citizens, convicted felons, etc.) or in active abstention because the alternatives available are not to the taste or interests of specific segments of the electorate (e.g. Burnham 1982; McDonald and Popkin 2001). Finally, many people who do not turn out to vote probably do so because they have other priorities at or around election time that prevent their participation: working away from home, failing to register because they have moved (in countries that require this), or even not knowing how to cast a ballot or where to cast it. This sort of indifference can be seen as reflecting a view of electoral politics as not so much systemically problematic (as with individualist apathy), lacking in appealing alternatives (protest), or the result of suppressing votes but as simply a routinized activity like all others in a modern society in which some people are simply not seriously invested. This can be symptomatic of abstention from politics in general more than simply from participation in a given election (e.g. Galli 2012).

Turnout rates are most frequently reported as a single statistic for an entire democracy. Though convenient, the practice of using a single statistic to summarize the behavior of national electorates obscures the geographically contingent process of voting. Furthermore, it encourages the use of sweeping generalizations to characterize political attitudes and behaviors as universal across a democracy. Recognizing that electoral participation varies across a democracy, explanations for voter turnout may benefit from approaches that are more geographically sensitive. Rather than devising a single, universal model of turnout for an entire polity, developing domain-specific models to examine turnout within particular geographic contexts, such as regions, states or provinces may yield improved insights into the decision to vote or not to vote, and more generally, into judgments about the universal and the particular.

Though parties and candidates devise national platforms on many issues, campaign strategies often reflect and make use of regional themes or local points of reference that may mobilize citizens to vote or even to abstain. Regional economic expansion or contraction, the introduction and implementation of a national welfare or educational program in certain places, or the triumphs and travails of a family or individual are often used as contextual backdrops during national campaigns, and serve to bring the issues, parties and candidates of the day home to the voter, where the voter will hopefully vote. Identifying and evaluating such sub-national contexts, and examining and comparing the correlates of electoral participation within and between such contexts, helps to clarify the socio-geographic processes underlying electoral participation (Diamanti 2012).

This geographic analysis of voter turnout focuses upon contemporary Italian democracy. Italy is a compelling case in which to examine the relationship between demography and democracy for several reasons. First, Italy is renowned for its historically high voter turnout rates. In the eighteen national elections since 1948, well over 80 % of voters turned out in each and every political contest. Second, as alluded to previously, Italy has one of the highest proportions of citizens aged 65 and over in the world. Coupled with one of the world's lowest birthrates, it is projected that Italy will experience a significant population decline in the decades to come. Third, Italy is often considered to be divided into two distinct geographic regions: a prosperous, civic-minded and European-oriented north versus an underdeveloped, peripheral, stagnant and corrupt south. Sub-divisions of this twofold division are also very common, as we discuss later. Fourth, the economic crisis of 2008 ushered in a remarkable period of social, political and economic turmoil across the Eurozone, and resulted in several austerity measures, some of which have been particularly acute for Italy and many Italians. Finally, the 2013 general election marked low-points in terms of voter turnout, and more generally, public attitudes towards Italian politics and politicians. Recognizing how such factors influence political participation differently in different places across Italy will shed light on the age-turnout nexus, and highlight the value of incorporating spatial concepts and theories into demographic-based accounts of politics.

9.2 Turnout, Age and Place in Italy

Electoral studies struggle with, "the Scylla of hasty overgeneralization and the Charybdis of myopic attention to local and national peculiarities" Rokkan (1966, 265). In many respects, the theories and methods of electoral studies still tend to predispose research into making either gross overgeneralizations or to providing local and particular anecdotes about voting behavior, with little middle-ground in between. Consequently, explanations of voter turnout usually fall into one of two categories (Niemi and Weisberg 1993). The first category considers the act of voting to be primarily a function of individual and psychological factors such as a voter's degree of party identification or attitudes on various social or economic issues. The second category considers group and sociological factors such as political mobilization and group membership to play an important role in the decision to vote.

With regard to the first category, there is a long tradition in political science and other disciplines to identify and evaluate the individual correlates of electoral participation (Franklin 1996; Lijphart 1997). For instance, one of the most cited predictors of turnout is level of education. People who are more educated tend to vote in higher proportions than those with less education (Caramani 1996; Denver and Hands 2004; Wolfinger and Rosenstone 1980). Based on responses from large scale pre- and post-election surveys, the turnout-education nexus is premised upon the argument that individuals with higher levels of education are more likely to follow the news, express a high degree of civic engagement, be interested in politics and political issues, and are thus more likely to vote (Wattenberg 2002). There is also the expectation that the middle-class and the wealthy are more likely to turn out than those with lower incomes, and that younger and older citizens are less likely to cast ballots in elections.

Contrasting such individual approaches to voter behavior are perspectives that consider group factors and socialization to be important determinants of turnout. Though an individual's income, religiosity, age, and level of education may indeed be related to political attitudes and behaviors, group effects and social interaction are posited to actually shape and an influence them. For instance, political campaigns often target certain groups (e.g., unions, retirees, factory-workers) in an effort to generate electoral support. Should such mobilization efforts resonate with group members, they may be more likely to vote and to encourage other group members to vote as well. Such effects are not limited to election campaigns (Goldstein and Ridout 2002; Johnston 1986), but include canvassing by parties and candidates (Krassa 1988), memberships to clubs and associations (Putnam 1993), as well as conversations between individuals (Johnston and Pattie 2000; Leighley 1990).

It is unlikely that the act of voting is determined by individual or group effects alone. Treating each approach to of voter turnout as mutually exclusive provides theoretical clarity, but also obscures important insights into the motivations behind voting. The middle-range approach that we use recognizes that the vote is an individual act but is also a function of group socialization. The concept of age as it relates to politics is simultaneously an individual trait and a social determinant. Age itself is used as a threshold both to vote and to hold political office, and cohorts, generations and group membership often use age to define membership. While age in itself can provide insights into political attitudes and behaviors, the types and intensity of political socialization within cohorts or generations can also be a function of age (e.g., the use of social media was crucial to influence young voters in recent US elections). In this respect, a middle-range approach that appreciates the particular and more general influences of age on turnout may yield important insights into voting.

As suggested above, age matters in relation to politics because it reflects both life cycle changes and generational effects. As people age their identities and interest shift. What was once important can become less so. Concerns about income security, for example, become more important than increasing the level of income. Access to reliable health care facilities becomes more important than the quality of local schools. At the same time, people come to political maturity in vastly different historical circumstances. For example, in Italy, those who first voted in the 1948 elections had just lived through 20 years of Fascism and were faced by two "families" of political parties on the left and right. Those first voting in the 1990s were introduced to electoral politics after the end of the cold war, the removal of compulsory voting, the collapse of a truly left-right dimension to party ideologies and the arrival on the electoral scene of the media baron and political salesman Silvio Berlusconi. The political memories of these generations are therefore fundamentally different. If the "new" politics fails to engage in some way with the older vocabularies and repertoires of political engagement, as well as with new interests and worries, then overall participation in electoral politics may also begin to tail off.

The experiencing of aging in both registers – life cycle and generational – is likely to be strongly mediated by the biographies acquired in living in the distinctive places out of which "Italy" is made on a daily and annual basis. This in turn can be expected to affect attitudes towards participation in elections aggregated across local and regional populations in terms of apathy, protest, and indifference. As noted previously, in Italy as a whole older voters tend to have the overall highest propensity to abstention. Though this is particularly marked among those born before 1926 (from 1985 through 2001), successive generations show a similar trend with the relatively oldest tending to the greatest propensity to non-voting (Tuorto 2006, 91). Some of this can be put down to "being too old to vote" in the sense of losing complete interest in politics or being unable to participate physically in the act of voting (because of the increasing number of extremely old people). The particularly large drop-off relative to other generations of that of 1926 can also be ascribed to that generation's socialization into that world of political parties after the Second World War that collapsed in the early 1990s and that has not been replaced by anything recognizably as stable.

But how does the relationship between age and turnout vary empirically across Italy? Differences between places are not mere "noise" disturbing presumably general relationships at the national level. Not only do proportions of the electorate in different age groups vary from place-to-place, primarily as a result of differential rates of in-and–out migration as well as historic differences in rates of family size and individual mortality, but how these demographic differences translate into differences in political behavior are determined by local economic and social conditions, cultural traditions, and relationships to parties and the political system more generally. In other words, being 'retired', 'young' or of 'middle-age' provides a demographic identity at one level, but the meaning – and the attitudes and behaviors associated with them – arguably vary from place to place.

Thus to understand the nature of the relationship between age and turnout necessitates examining the ways in which the two vary and co-vary spatially. Any generalizations at the national level must therefore be drawn from establishing the character of the distributions and relationships from within local and regional contexts. Figure 9.1 illustrates this point by plotting the national voter turnout rate in Italy since 1948, and the turnout rates for three of the 100+ Italian provinces, namely, Milan in the north, Rome in the center, and Palermo in the south of the peninsula, since 1987. Similarly, Fig. 9.2 plots the population pyramids for: (a) Italy; (b) the administrative region of Lombardy, home to Milan; (c) the region of Lazio where Rome is located; and, (d) the island region of Sicily where Palermo is situated.

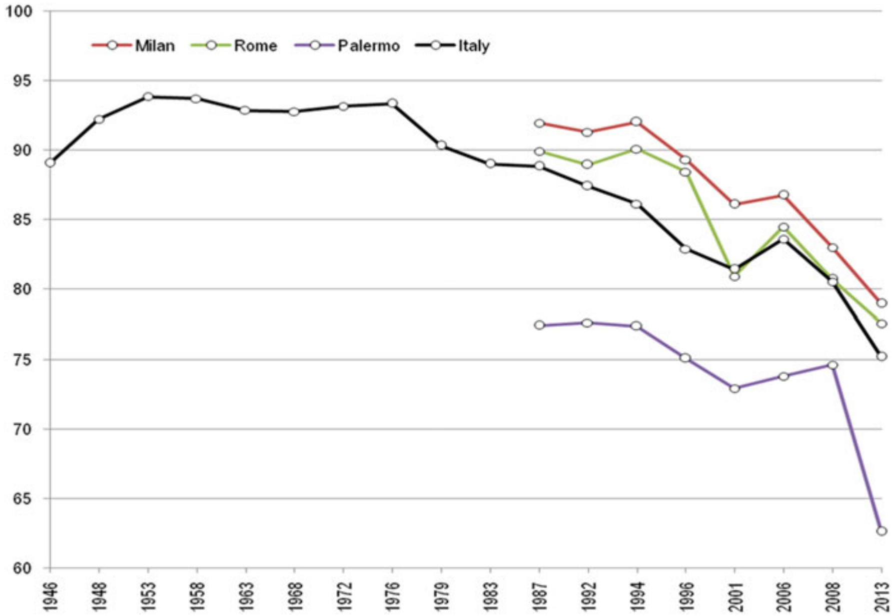

Fig. 9.1 Voter turnout in Italy and selected provinces, 1946–2013. Provincial turnout data not available before 1987 (Source: Ministero dell'Interno)

Declining voter turnout in Italy is apparent as early as 1979, and is marked by a sharp downward trend since the 2006 national election. This general trend is reflected in the turnout rates returned in Milan, Rome and Palermo, but the local changes and trends are not consistent between the provinces. For instance, voter turnout actually increases in Palermo between 2006 and 2008. Such differences illustrated in Fig. 9.1 suggest that there are probably more fundamental or structural differences in voter behavior between these and other provinces, which are hidden by the national average. A comparison of the national and regional population pyramids also reveals differences in population numbers and the age structures of these three regions. Not only is the population of Sicily much smaller than that of Lazio (Rome) and Lombardy (Milan), but the profile of its age structure differs as well. In particular, the size of population groups tend to be more equal in size in Sicily than in Lombardy or Lazio, which display notable middle-age bulges (i.e., 30–60 years). The key takeaway from this set of figures is that national aggregates and averages (e.g., turnout rates and population structures) both comprise and hide significant subnational differences. Determining whether such subnational variations in turnout and voting are substantive is the focus of the remainder of this chapter.

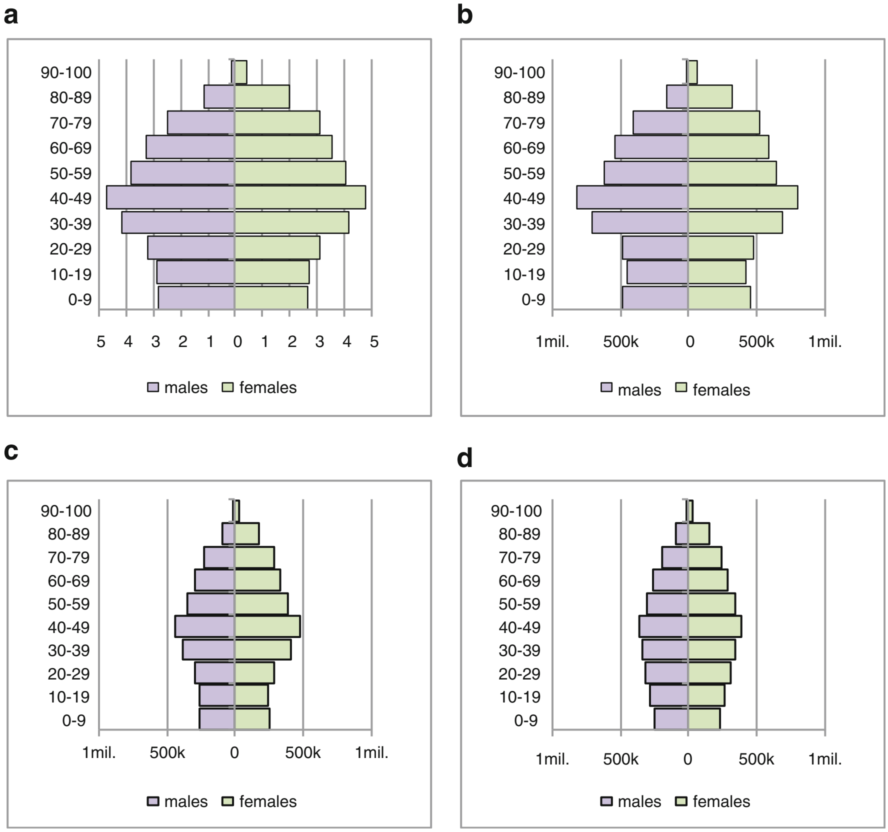

Fig. 9.2 Population pyramids for: (**a**) Italy (in millions); (**b**) Lombary-Milan; (**c**) Lazio-Rome; (**d**) Sicily-Palermo (Source: ISTAT)

9.3 Data & Methods

Drawing from the previous discussion, we specify two general hypotheses concerning the linkage between age and voter turnout to guide the following analysis. First, we expect voter turnout to vary significantly across Italy, and to be significantly lower in the south than in the north. This regional difference in turnout is well-known, but the 2013 Italian general election presents itself as an important opportunity to reassess this divide, and possible changes to it, because it marked the lowest level of voter turnout in the history of modern Italian democracy. Second, we expect that the relationships between the same age groups situated in different places in Italy and voter turnout will not be consistent across the country. We contend that age and demographics are mediated differently in different places, and subsequently, that political attitudes and behavior are geographically contingent.

To assess the geographic dimensions of the age-turnout nexus, we use demographic and election data for the 103 Italian provinces and spatial regression techniques. Demographic and other socio-economic data were obtained from the Italian National Statistical Agency (ISTAT) and election data were provided by the Cattaneo Institute located in Bologna, Italy. Given our focus on the spatial demographics of voter turnout, the following analysis is informed and guided by geospatial analysis and spatial econometric techniques (Anselin 1988). Since most demographic and election data are compiled and aggregated on a geographic basis, for instance, by census tracts or election precincts, any analyses using such data must recognize the biases and limits inherent to such data. Formal spatial analytic techniques not only identify issues such as spatial autocorrelation, or non-random clustering, across data sets, but also offer methods to incorporate or control for such effects in regression models (e.g., Anselin 1988, 1995; O'Loughlin et al. 1994; Shin and Agnew 2011). This geographically sensitive approach is especially appropriate for this investigation into middle-range perspectives on demography and democracy because it simultaneously recognizes the limits of theoretical overgeneralization and myopic attention to particularities.

Figure 9.3 maps the key variables of interest within the scope of this analysis of contemporary Italian voter turnout, namely, voter turnout in 2013 and changes in turnout since 2008. We use a very common geographic division between the 'north' and 'south' of Italy in this analysis and include it as a dividing line on the maps. Though a discussion of Italy's 'southern question' is beyond the scope of this chapter, this geographic distinction has long been a concern for many Italians, policymakers, and academics (e.g., Banfield 1958; Trigilia 1992; Davis 1996). Moreover, this geographic division of the peninsula is formalized through the reporting of statistics for the north and south by the Italian national statistical agency, ISTAT. It is used here to demonstrate the utility of middle-range perspectives that draw insight from both the national and local levels, and the geographic variations between and within them.

The top two maps in Fig. 9.3 show levels of turnout in the 2013 general election and changes in turnout since the last 2008 general election. Despite the fact that turnout in 2013 was the lowest ever recorded in post-World War II Italian democracy, compared to other democracies it remains relatively high, especially across the north. Turnout increased in only one of the 103 provinces (i.e., Campobasso) since the 2008 general election, and electoral participation dropped off considerably more in the south than in the north.

The bottom two maps highlight the spatial clustering of similarly high and low levels of turnout and turnout change, respectively. A local indicator of spatial association (LISA) index is calculated for each province which measures the degree to which levels of turnout (and turnout change) in one province are correlated with those that are found in neighboring provinces (see Anselin 1995; Shin and Passarelli 2012). Statistically significant values for the LISA index are mapped, thus revealing that higher levels of turnout are concentrated in north central Italy, and clusters of low turnout are found in the south and islands of the peninsula. Clusters of recent turnout change are not as extensive as those for turnout, but the

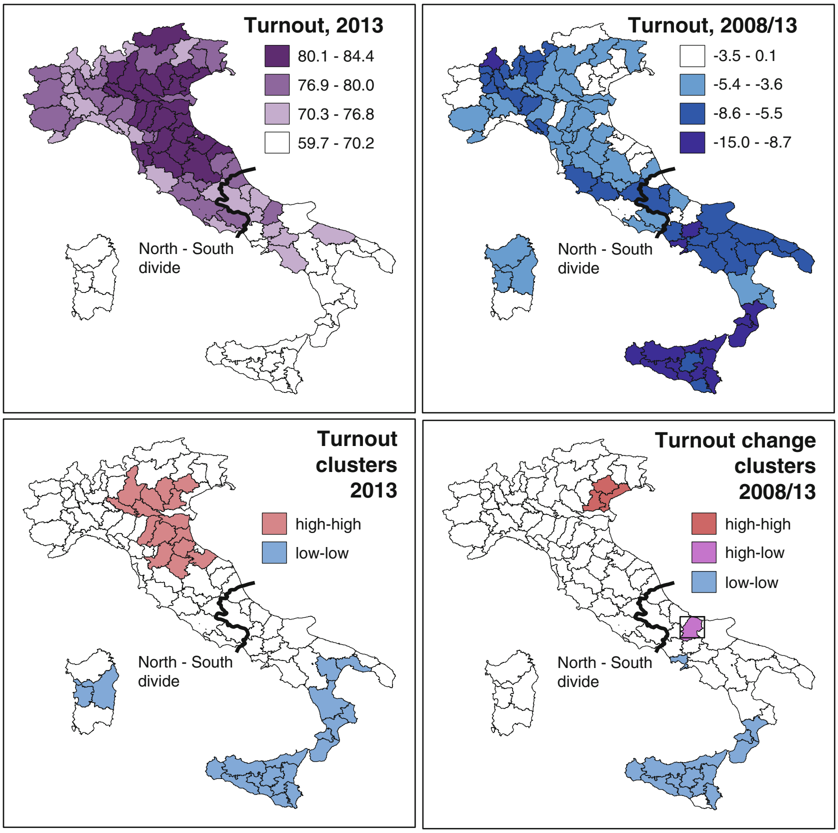

Fig. 9.3 Geographic distribution and clustering of 2013 voter turnout and turnout change, 2008–2013

map again shows geographic differences between the north and south, and the clustering of low turnout and large declines in turnout across Sicily. Note that for the turnout change cluster map, 'low-low' refers to the clustering of similarly large declines in turnout, and 'high-high' refers to clustering of comparatively small changes in turnout. The province of Campobasso in the south is considered a spatial outlier because it is the only province where turnout increased, by a mere +0.07 %, and was surrounded by negative change values.

In light of the notable geographical clustering and variations in electoral participation and turnout change, we contend that a single national model of voter turnout is grossly insufficient. Focusing on local particularities, such as Campobasso, would also be unsatisfactory. We proceed by estimating a national baseline ordinary least squares model (OLS: $y = a + bX + e$), and a set of spatial regression models for the north, the south, and all of Italy for comparison. Based on the results from the above spatial analyses and other diagnostic tests (unreported), our spatial econometric approach incorporates a spatial lag term (i.e., $y = Wy + a + bX = u$).

For each and every observation, the spatial lag, *wy*, contains the weighted average of all neighboring observations, and is often used to capture the effects of proximity in spatial econometric modeling (see Anselin 1988). Like the LISA index, the spatial lag captures the degree to which turnout (or changes in turnout) in one province are correlated with turnout in neighboring provinces.

The decision to estimate separate sets of models for both the north and south stems from our position that places mediate, condition and shape the very definitions and activities of individuals and social groups. The inclusion of a regional dummy variable (e.g., south = 1) is insufficient because the effects of region and place are not additive but contextual. For example, we contend that being an educated, white collar, manager in the north around Milan is fundamentally different from being an educated, white collar, manager in Palermo, Sicily, which in turn may lead to differentiated political attitudes and behaviors. In this respect, we reject the notion of the idealized median national voter. Moreover, differences in the very nature of the spatial relationships between places (i.e., provinces), as captured by the spatial lag term, can be identified and assessed.

The dependent variables used in the following analysis are: (a) provincial levels of 2013 voter participation; and, (b) provincial level changes in voter participation since 2008. Drawing from the large body of literature on electoral participation across contemporary democracies (Wattenburg 2002; Gimpel et al. 2004), we included levels of GDP (1,000s of Euros), university graduates and the unemployment rate as covariates. Based on previous studies and understandings of voter behavior, the first two of these variables are expected to be positively associated with turnout, but a negative relationship is expected to emerge between the provincial unemployment rate and turnout.

To capture and assess the relationship between age and turnout, we use provincial age distributions. Specifically, we calculated the proportion of the 2012 provincial population that belonged to a particular cohort. Cohorts are characterized and often defined by external events and shared experiences, such as the fall of the Berlin Wall, and tend to be shorter than generations that are typically defined by years of birth. At the same time, cohorts can provide glimpses into life cycle effects on turnout discussed earlier. Drawing from previous work on political generations in Italy (Corbetta and Parisi 1994; Corbetta 2002; Caramani 1996; Tuorto 2006, 2010; Legnante and Segatti 2009), seven different political cohorts since 1945 are defined and reported in Table 9.1.

For each cohort, we report the range of years of birth, the years in which the cohort became eligible to vote, the range of ages in the cohort in 2012, key political events for the cohort, and the cohort's expected relationship with levels of voter turnout. Expectations for the relationships between cohort and turnout are drawn from the body of work identified above. Generally, voters on the margins or periphery of socio-political life (e.g., the old, the young, under-educated, poor) are less likely to vote, so we expect a negative relationship to exist between turnout and the two oldest (I, II) and the youngest cohorts (VII). The notion of marginality is arguably reinforced in the south of Italy, and we expect there to be notable north-south differences in the association between these peripheral cohorts. We expect

Table 9.1 Political cohorts in Italy and expected relationship to voter turnout

	Year of birth	Eligible to vote in...	Age in 2012	Defining events for cohort	Expected relationship with turnout
I	Pre-1927	<1945	>85	Fascism, World War II	−
II	1927–1946	1945–1964	66–85	Postwar, cold war, Italian economic miracle	−
III	1947–1955	1965–1973	57–65	1968, Prague spring	+
IV	1956–1965	1974–1983	47–56	National unity government, Red Brigades	+
V	1966–1974	1984–1992	38–46	*Pentapartito*, fall of Berlin Wall, *Tangentopoli (bribesville), Mani pulite (clean hands)* scandals	+
VI	1975–1985	1993–2003	27–37	Berlusconi I, Gulf Wars, 9/11	+
VII	1986–1994	2004–2012	18–26	Berlusconi II, wars in Iraq and Afghanistan, global recession	−

positive relationships between the middle age cohorts (III–VI) and turnout because these groups are typically the most politically informed and engaged. Our expectations for the associations between cohorts and change in turnout reflect those for turnout. Though our expectations are 'generationally' informed and derived, the use of shorter cohorts in some periods provides demographic granularity that may permit further insights into the linkages between age and turnout.

9.4 Results

Table 9.2 reports the estimates from the baseline OLS and spatial regression models. For clarity, aside from the national baseline model, only statistically significant estimates from the best fitting, most parsimonious models are reported. Looking first at the results from the baseline OLS model, three variables are significantly associated with voter turnout in 2013: provincial levels of unemployment; the pre-World War II cohort (I); and, the Berlusconi I cohort (VI). Though this model explains a considerable amount of overall variance (86 %), diagnostic tests (unreported) indicate that the model suffers from a range of issues such as multicollinearity, a non-normal error term, and spatial autocorrelation in the error term. Failure to address these specification issues can result in biased and inefficient parameter estimates, inaccurate significance tests, and most importantly, incorrect conclusions. To overcome these issues, and as noted previously, we proceed with spatial econometric techniques, and in particular a spatial lag model, that is designed to incorporate spatial effects (see Anselin 1988; Ward and Gleditsch 2008).

Table 9.2 "Estimates from the baseline OLS and spatial lag models, estimates in **boldface** are significant at the 0.05 level

	National	National	National	North	North	South	South
	Turnout 2013	Turnout 2013	Turnout change	Turnout 2013	Turnout change	Turnout 2013	Turnout change
	Baseline OLS	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
Constant	59.167	14.456	−1.808	70.304	−0.641	43.596	−1.731
Spatial lag	–	**+0.403**	**+0.470**	**+0.409**	**+0.599**	**+0.387**	**+0.370**
GDP (1,000 s)	0.069		**+0.171**		**+0.115**		
Laureati	0.195				**0.352**		
Unemployed	**−0.566**	**−0.512**				**−0.396**	**−0.166**
I. Pre-WWII	**3.452**	**+1.672**				**+2.180**	
II. Cold War	−0.685						
III. 1968	−1.193			**−2.586**	**−0.566**		
IV. Unity	1.022						
V. Cold War II	0.273						
VI. Berlusconi I	**1.734**	**+1.940**		**+1.212**			
VII. Youth	−1.163			**−1.528**			
R-squared	0.861						
pseudo-R-squared		0.873	0.506	0.670	0.523	0.769	0.349
N	103	103	103	67	67	36	36

All estimates obtained from the GeoDA spatial analysis software package (Anselin et al. 2006)

Looking across all of the lag models in Table 9.2, for turnout and turnout change, the only consistently significant variable is that for the spatial lag term. In other words, levels of provincial turnout and turnout change are positively related to levels in neighboring provinces. A coefficient of +1.0 would indicate that turnout in a given province could effectively be predicted by the average level of turnout in adjacent provinces. The predictive strength of the turnout lag term remains relatively consistent for the entire country (lag 1), as well as for models restricted to the provinces in the north (lag 3) and south (lag 5). The lag term for turnout change is larger than for turnout alone in the national (lag 2 v. lag 1) and north (lag 4 v. lag 3) models, but is smaller in the south (lag 6 v. lag 5). These results indicate that spatial influences on recent levels of turnout are moderate in size and consistent across Italy and the north (lag 1 and lag 3); spatial effects have a greater impact on turnout change when considered nationally and in the north (lag 2 and lag 4); and, sub-national spatial effects are larger in the north than south (lag 3 v. lag 5, and lag 4 v. lag 6). Geography indeed matters with regard to voter turnout and changes in turnout, but its influence is spatially differentiated.

With regard to variables that are significant in the national, north and south models, it is interesting to note that no single cohort is significantly associated with turnout in both the north and south. For instance, the Berlusconi I (VI) cohort is significant in the national and north models, but not in the south. This cohort includes those voters who became eligible to vote at the same time that Silvio Berlusconi entered the Italian political scene. This was also the period where the old system of parties was dismantled and replaced by a new one that offered existing voters an entirely different menu of political choices (see Shin and Agnew 2002, 2008). Between 1993 and 2003, one of Berlusconi's key political allies, the regionalist Northern League party, built a platform around the secession of the north, and vociferously characterizing the south and southerners as chronically backwards and corrupt, among other things (see Agnew 1995; Diamanti 1996). The re-emergence of Berlusconi on the 2013 ballot, in the face of government austerity and economic recession, may have energized some members of this cohort to turnout, especially in areas of the north where he was popular previously.

The positive relationship between the pre-World War II (I) cohort and turnout is opposite of what was expected and is somewhat puzzling. One possible explanation for this result is that poor weather dampened turnout across much of northern Italy in the 2013 election, and may have exaggerated the effects of 'normal' turnout levels for this already marginalized cohort in the south. The negative relationship between what is called the '1968' (III) cohort and turnout is also opposite of what was expected. This may reflect the wide-spread disgust, anger and resentment towards government austerity, entitlement changes, political stalemate and economic stagnation of those quickly approaching retirement and expecting government pensions, and subsequently, their choice to stay away from the polls. Negative associations between turnout and levels of provincial unemployment are significant nationally and in the south, and between the youngest cohort (i.e., VII) in the north. Though this result was expected, the unemployment covariate in the south is also probably picking up this youngest cohort, as recent figures estimate youth unemployment (ages 16–24) in Italy to be over 34 %, and female youth unemployment in the south to exceed a staggering 50 % (ISTAT 2013). Finally, since 2008 it appears that demography has mattered relatively little to changes in voter turnout. This is likely to be a reflection of the near universal political disaffection resulting from economic stagnation and hardship that all Italians are experiencing.

The above results support the position that an intermediate, middle-range approach to understanding the age-turnout nexus is in fact useful. By leveraging the geographic structure of voting data, and the granularity of population data, domain-specific (i.e., national, north, south) models underscore the need for a more nuanced, middle range approach. Such approaches also permit interesting and valuable comparisons to be made, which can inform and extend the complementary use and understanding of both theory and method across spatial demography.

9.5 Conclusion

Long standing geographic differences in both turnout and turnout change persisted between the north and south in Italy's 2013 general election. Spatial effects, or the influence of the local provincial context, were significant but also displayed notable geographic variations. Similarly, the associations between cohorts and turnout (and turnout change) were not consistent across Italy. For instance, a positive relationship between the cohort of voters who have only known an Italian politics dominated by Silvio Berlusconi and turnout was detected, but this linkage appeared only in the north of Italy. Moreover, age profiles seem to matter little with regard to the precipitous decline in Italian electoral participation. It seems that the effects of the recent economic recession and ongoing political crises of Italy are broad and far-reaching. That said, it is precisely such events and circumstances that define geographically situated cohorts and generations, and that shape political attitudes and behaviors.

Appreciating and understanding the linkage between age and voting requires theoretical and methodological approaches that are sensitive to such global trends, sub-national patterns and local idiosyncrasies. Neither the local nor the national are privileged in mid-range approaches, but both are recognized as necessarily complementary. Moreover, spatial analytic and spatial econometric techniques that can reveal and evaluate clustering, regional variations and local dependencies highlight the need for such perspectives that draw from multiple scales of analysis. As this chapter illustrates, spatial demography is very well positioned to both benefit from and promote such middle-range perspectives and techniques.

As the number of elections around the world continues to rise, it remains to be seen whether or not electoral participation in Italy and other democracies will continue to decline. Some consider such declines in voter turnout a challenge to the overall state of democracy in the world, but others contend that they are of little consequence. Questions concerning the causes and consequences of voter turnout and abstention will continue to be important in the future, and answering such questions from a spatial demographic perspective will certainly extend current thinking on political participation in both new and established and democracies.

References

Agnew, J. (1995). The rhetoric of regionalism: The Northern League in Italian politics, 1983–94. *Transactions of the Institute of British Geographers, 20*, 156–172.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis, 38*, 5–22.

Banfield, E. (1958). *The moral basis for a backwards society*. Chicago: Free Press.

Burnham, W. D. (1982). *The current crisis in American politics*. Oxford: Oxford University Press.

Caramani, D. (1996). La partecipazione elettorale: gli effetti della competizione maggioritaria. *Rivista Italiana di Scienza Politica, 26*, 585–608.

Corbetta, P. (2002). Le generazioni politiche. In M. Caciagli & P. Corbetta (Eds.), *Le ragioni dell'elettore*. Bologna: Il Mulino.

Corbetta, P., & Parisi, A. (1994). Smobilitazione partitica e astensionismo elettorale. *Polis, 8*, 423–443.

Davis, J. (1996). Changing perspectives on Italy's 'Southern Problem'. In C. Levy (Ed.), *Italian regionalism* (pp. 53–68). Oxford: Berg.

Denver, D., & Hands, G. (2004). Exploring variations in turnout: Constituencies and wards in the Scottish Parliament Elections of 1999 and 2003. *The British Journal of Politics and International Relations, 6*, 527–542.

Diamanti, I. (1996). *Il male del Nord*. Rome: Donzelli.

Diamanti, I. (2012). *Gramsci, Manzoni e mia suocera. Quando gli esperti sbagliano le previsioni politiche*. Bologna: Il Mulino.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper and Row.

Franklin, M. N. (1996). Electoral participation. In L. LeDuc, R. Niemi, & P. Norris (Eds.), *Comparing democracies: Elections and voting in global perspective* (pp. 216–235). London: Sage.

Fruncillo, D. (2004). *Urna del silenzio. L'astensionismo elettorale in Italia*. Rome: Ediesse.

Galli, C. (2012, May 24). Astensionismo. Se il partito del non voto diventa maggioranza. *La Repubblica*.

Gimpel, J., Dyck, J., & Shaw, D. (2004). Registrants, voters and turnout variability across neighborhoods. *Political Behavior, 26*, 343–375.

Goldstein, K., & Ridout, T. (2002). The politics of participation: Mobilization and turnout over time. *Political Behavior, 24*, 3–29.

ISTAT. (2013). *Occupati e disoccupati: dati recostruiti dal 1977*. http://www.istat.it/it/archivio/88827

Johnston, R. J. (1986). Places, campaigns and votes. *Political Geography Quarterly, 5*, S105–S117.

Johnston, R. J., & Pattie, C. (2000). People who talk together vote together?: An exploration into context effects in Great Britain. *Annals of the Association of American Geographers, 90*, 85–111.

Krassa, M. (1988). Context and the canvass: The mechanisms of interaction. *Political Behavior, 10*, 233–246.

Legnante, G., & Segatti, P. (2009). Intermittent abstentionism and multi-level mobilization in Italy. *Modern Italy, 14*, 167–181.
Leighley, J. (1990). Social interaction and contextual influences on political participation. *American Politics Quarterly, 18*, 459–475.
Lijphart, A. (1997). Unequal participation: Democracy's unresolved dilemma. *American Political Science Review, 91*, 1–14.
McDonald, M. P., & Popkin, S. (2001). The myth of the vanishing voter. *American Political Science Review, 95*, 963–974.
Niemi, R., & Weisberg, H. (1993). *Controversies in voting behavior* (3rd ed.). Washington, DC: Congressional Quarterly Press.
O'Loughlin, J., Flint, C., & Anselin, L. (1994). The geography of the Nazi vote: Context, confession and class in the Reichstag election of 1930. *Annals of the Association of American Geographers, 84*, 351–380.
Putnam, R. D. (1993). *Making democracy work*. Princeton: Princeton University Press.
Riker, W., & Ordeshook, P. (1968). A theory of the calculus of voting. *American Political Science Review, 62*(1), 25–42.
Rokkan, S. (1966). Electoral mobilization, party competition and national integration. In J. LaPalombara & J. Weiner (Eds.), *Political parties and political development*. Princeton: Princeton University Press.
Shin, M., & Agnew, J. (2002). The geography of party replacement in Italy, 1987–1996. *Political Geography, 21*, 221–242.
Shin, M., & Agnew, J. (2008). *Berlusconi's Italy*. Philadelphia: Temple.
Shin, M., & Agnew, J. (2011). Spatial regression for electoral studies: The case of the Italian Lega Nord. In B. Warf & J. Leib (Eds.), *Geography revitalizing electoral* (pp. 59–74). Surrey: Ashgate.
Shin, M., & Passarelli, G. (2012). The Northern League in national, European and regional elections: A spatial analysis. *Polis, 26*, 355–370.
Trigilia, C. (1992). *Svillupo senza autonomia*. Bologna: Il Mulino.
Tuorto, D. (2006). *Apatia o protesta? L'astensionismo elettorale in Italia*. Bologna: Il Mulino.
Tuorto, D. (2010). La partecipazione al voto. In P. Bellucci & P. Segatti (Eds.), *Votare in Italia, 1968–2008* (pp. 53–79). Bologna: Il Mulino.
Ward, M., & Gleditsch, K. (2008). *Spatial regression models*. Thousand Oaks: Sage.
Wattenberg, M. (2002). *Where have all the voters gone?* Cambridge: Harvard University Press.
Wolfinger, R., & Rosenstone, S. (1980). *Who votes?* New Haven: Yale University Press.

Chapter 10
A Spatial Decomposition of County Population Growth in the United States: Population Redistribution in the Rural-to-Urban Continuum, 1980–2010

Jeremy R. Porter and Frank M. Howell

10.1 Introduction

A significant theme in demographic studies has been the population redistribution patterns among metropolitan centers, non-metropolitan areas surrounding them, and the so-called hinterlands beyond. Demographers, in particular, have spent a great deal of effort toward understanding the trends, patterns, and reasons for population dynamics in rural and urban areas of the U.S. (Brown and Wardwell 1980; Frey 1987). For example, the attention given to the rural population turn-around during the 1970s (Brown and Wardwell 1980) and how it tended to subsequently turn "back around" to decline in the 1980s (Frey 1993; Frey and Speare 1992), only to reverse itself somewhat again during the early 1990s (Johnson and Beale 1994), gives witness to the importance of rural–urban population dynamics by demographers and others. We also note that virtually all of this research has used the metropolitan vs. non-metropolitan classification scheme, whether the data source is the Current Population Survey or county population

Direct all communications to: Frank M. Howell, frankmhowell@gmail.com or Jeremy R. Porter, jporter@brooklyn.cuny.edu. Several students in the Spatial Analysis of Social Data course taught by the first author in the Department of Sociology, Anthropology and Social Work at Mississippi State University made helpful suggestions contributing to this chapter. Any errors of fact or interpretation, however, are those of the authors.

J.R. Porter (✉)
Brooklyn College and Graduate Center, City University of New York, Brooklyn, NY, USA
e-mail: jporter@brooklyn.cuny.edu

F.M. Howell
Department of Sociology, Mississippi State University, 416 Bay Cove,
Ridgeland, MS 39157, USA
e-mail: frankmhowell@gmail.com

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_10

data (Lichter 1993). This choice, however convenient, has implications for the findings of this body of research, as noted below.

The causes of these patterns of rural-urban population redistribution have been described by three competing perspectives (Frey and Speare 1992; Lichter 1993). The *period-effects*, *regional restructuring*, and *deconcentration* (Frey and Speare 1992) perspectives are slightly overlapping, yet complementary, views but they have not yielded fully adequate understandings of these population dynamics. For instance, Johnson and Beale (1994: 665) state that these theoretical explanations of future population change in rural areas are "perilous" and are "likely to be more volatile than in the past," after studying change using population estimates through the first portion of the 1990s. On the other hand, Frey and Speare (1992: 144) suggest that a "continued preference among residents to live and work within large metropolitan areas" is evident from the last half of the 1980s. While these conclusions are not completely contradictory, they illustrate the clear need for additional theoretical development. Given these contrasting perspectives, Johnson and Beale (1994: 666) also suggest that "careful monitoring of future non-metro demographic trends" is vital for informing both theory and policy-making in the United States.

In a 1992 follow-up, Frey examined the metro/non-metro population trends in the 1980s. In this article he introduces a third perspective called the Period Explanations Perspective, which saw the 1970 shift as a distortion in the 'normal' traditional trend of urbanization. This perspective believed that the population decline in the metro areas of the 1970s was directly related to a number of unique economic and demographic circumstances (Frey 1992). First, there was the deindustrialization and energy crisis at the time that forced many out of the Northeast and into the South and West where the energy crisis had stimulated natural resource exploration. In addition, the large baby boomer cohort was coming of age and increased many small college town populations during this time period. They then were forced to the South and West as they were unable to find jobs in the oversaturated labor market of the North (Frey 1992). Again, these factors are seen as directly relating to the distortion in the traditional population trend and therefore the trend should regress to its normal trajectory as these unique issues disappear.

At the same time, there has been a considerable long-term debate over the conceptual definition of rural locales and, more recently, concern over the definition of metropolitan statistical areas themselves (Dahmann and Fitzsimmons 1995; Federal Register 1999). The struggle for conceptual refinements of the rural-urban continuum is relevant for the population redistribution phenomenon. As Lichter (1993: 19–20) put it, "What do[es]. . .population redistribution mean in an increasingly urban society? Current redistribution and migration trends clearly challenge us to rethink the conceptual and methodological tools at our disposal. One consequence is that the significance of population redistribution research may increasingly reside in analyses of population shifts *within* rather than *between* conventional units of analysis." A reliance on county-level data, for instance, to study population redistribution represents a clear example of this criticism: the county (and its equivalents) may be too internally heterogeneous to adequately capture these types of population shifts. There is thus a need to continue to monitor population redistribution patterns in the U.S. but with approaches sensitive to capturing current types of dynamics in the rural-urban continuum.

With the release of Census 2010 population data at various levels of geography, coupled with alternative concepts and methods, we extend previous research on population redistribution in the U.S. through the three full decades of the 1980s, 1990s, and 2000s. Our study uses a multi-level geography design which allows sub-county population data to be used to characterize the county. We offer a new sub-county geography, the *non-place territory* (see also Porter 2010 and Porter 2011) to complement incorporated (and Census-designated) places, as a step toward reducing the internal heterogeneity of counties as the unit-of-analysis. We combine this new geography with a method to spatially decompose county and place-level data using GIS procedures. These methods also allow for the visualization of population redistribution dynamics over the three decades of 1980–2010.

10.1.1 Purpose

The objectives of this study are to offer one new approach to study population distribution dynamics beneath the county-level. Our approach enhances the applicability of place-level geography by identifying: (a) place vs. non-place territory population concentrations during the past three decades, and (b) where county population growth has been driven by non-place territories (i.e., counties where non-place territory growth exceeds place-based growth). Building on the approaches of Tita and Cohen (1998), Howell (2004), Porter (2010, 2010) and others, it is expected that this study can effectively use a form of the local spatial clustering statistic (LISA) in order to identify locales of significant population loss and growth. Howell (2004) proposed using the within-county share over time to assess patterns of concentration vs. deconcentration. Within a larger county context, this change in share of that population can further highlight significant patterns of population deconcentration and concentration at the place level but *within* counties. This aspect of small-area population dynamics has not been examined in the published literature.

10.2 Relevant Literature

Two main themes in the extant literature are briefly reviewed, those involving non-metropolitan population change and those defining rural areas in the United States, followed by a delineation of the *non-place territory* concept.

10.2.1 Non-metropolitan Population Change

There has been much work published by demographers on population change in non-metropolitan areas of the U.S., especially in relation to their proximity to MSA's and factors that drive such change (Brown and Wardwell 1980; Lichter 1993; Brown and Zuiches 1993). This line of research has examined various perspectives regarding rural population change: *historical period effects* (Johnson 1989; Lichter 1993; Frey 1993), *deconcentration* (Vining and Strauss 1977; Lichter and Fuguitt 1982), and *regional economic restructuring* (Frey 1987; Kasarda and Irwin 1991). Most of this work has used county-level data or, in some cases, micro-data from the Current Population Survey (CPS) data program, to make generalizations about rural populations defined as rural by non-metropolitan standards.

Throughout the latter half of the nineteenth century demographers and social scientists alike have given a good deal of attention to better understanding and defining the rural/urban dichotomy. Perhaps one of the most often-studied phenomena associated with this complex strain of research is that of suburbanization and its impact on the affected local economy, ecology, and geography of an area. More often than not, the consensus is that the process of suburbanization has made the lines demarcating the lifestyles associated with urban and rural much less obvious. This intertwining of rural and urban has brought traditionally rural activities, such as those associated with agriculture, to what are now defined as metro areas (Thomas and Howell 2003). Likewise, the existence of a number of traditionally urban amenities, such as advanced communication and transportation resources, are now readily available in many areas defined as non-metro (Brown and Zuiches 1993). Facilitating these changes are a number of factors, two of which are perhaps more important than the others.

First, the migration tendencies of individuals have played a large role in the dispersal of not only people but also ideas. During the 1970s there was turnaround trend towards a deconcentration of the population to non-metro and rural areas. This slowed and slightly reversed through the early 1980s but remained the overall trend (Lichter 1992; Frey 1992). Many times this deconcentration is referred to as suburbanization or sprawl because of the associated housing and development booms that often take place in response to the shifting population. Advances in communication and transportation were important in that they allowed individuals to perform the same tasks without the necessity of the same spatial proximity. However, proximity does matter to a degree as not all rural areas saw an increase during this time of deconcentration as regional and proximity biases were still present (Isserman 2001). Those counties which tended to grow were usually adjacent to metro counties and the fastest growing were usually located in the south or west regions of the U.S (Lichter 1992).

The second factor, which has lent itself to the "restructuring" of what we think of as urban or rural, literally has restructured what we think of as urban and rural. The Office of Management and Budget (OMB) has itself facilitated the "rurban" phenomena by tweaking the definition of what is metro, the term most often used to separate urban from rural. This “re-definition”, coupled with the propensity of the population to deconcentrate has allowed for a number of previously non-metro counties on the fringes of Metropolitan Statistical Areas (MSAs) to be considered part of the MSA based on requirements of social and economic integration (basically commuting patterns).

In a sense, rural America was disappearing into metro America in terms of classification attributes for the purpose of the census or OMB, however in reality it may simply be traditional rural America masked as metro America (Isserman 2001). It can be assumed then that much of what is thought of as rural America will be present in what we today consider urban or metro America. This would include features such as the traditional small-town lifestyle and associated activities, one of the most dominant and time enduring being the traditional stronghold of agriculture or farming, which is thought of most often-taking place “out in the country”.

As Lichter’s much earlier review of this line of research has suggested, “There have been no clear winners in this debate, in fact, these perspectives often fail to provide mutually exclusive predictions. The reality is that current migration trends continue to reflect both concentrating and deconcentrating tendencies ... [these] theories .. provide a useful backdrop to the central question of changing spatial inequality” (1993: 34). Accompanying the early population shifts in the 1970s were a number of theories on why such trends were developing and what external factors were facilitating their development. Frey introduces two such theories in a 1987 article aimed at examining the deconcentration phenomenon.

The first theory introduced by Frey is the *Regional Restructuring Perspective*, which believed that the largest of the metropolitan areas would continue to grow as it served as the “command base”. These areas included cities like New York, Chicago, and Los Angeles. Based on this theory it was the smaller and mid-sized metro areas that were responsible for the 1970s population shift towards non-metro and rural areas. This perspective was grounded in the belief that as part of the newly developing global economy a kind of functional hierarchy would develop in which the largest metropolitan centers would continue to grow as they would serve as the headquarters and centers of major operation for transnational businesses. Likewise, the smaller and mid-sized metro areas would lose population as they transitioned from local industry sectors deeply rooted in out-dated industry to a more global economic service approach (Frey 1987).

As a result metro areas like Detroit, which were deeply rooted in the automobile industry, would be expected to lose population as the city transitioned to more of a service oriented metro area. According to Frey, the out-migration of these areas then would head towards smaller non-metro and rural areas in which specialized centers and subordinate centers would develop to support the headquarters, including smaller industry and manufacturing plants. These centers could develop in these non-metro regions because they, unlike the smaller and mid-sized metro areas, were not deeply rooted in specialized industry and therefore could easily build plants and centers to support the new industry. Other pull factors for non-metro/rural areas were cheaper expenses, including labor, land, and taxes.

The second and contrasting theory put forth by Frey was the *Deconcentration Perspective*, which stated that there would be a gradual but sustained depopulation of larger metro areas (1987). This perspective placed much less importance on the restructuring of the newly developing global economy and more emphasis on the technological advancements and human preferences. These advancements allowed workers and employers to follow, what Frey called, their natural preferences towards lower density residential and workplace locations, with lower crime rates, and better education districts; this led to what is sometimes called the "rural renaissance". This perspective suggests that new production locations will be picked increasingly based on residential location preferences. Clearly stated, as technological advancements allow the workers and consumers to move further from the densely populated metro areas the employers will follow as the competition for well-educated, skilled, and professional personnel rises (Frey 1987).

As a result the two perspectives expect and predict different growth tendencies in large metro areas as the deconcentration perspective expects all metro areas to sustain population reduction and the Regional Restructuring Perspective predicts metro areas would grow or not grow in a polarized fashion based on their size and ability to support and serve as the "command center" or headquarters of newly developing global and domestic businesses. Ultimately Frey's analysis came to accept the Deconcentration Perspective as his examination showed two developing trends. First, the post 1970 migration pattern showed depopulation in the largest metro areas of the North and secondly non-metro areas primarily in the South showed the largest net gain in population (1987). So here you not only had a metro to non-metro shift in population gains but you also had a North to South population shifts, both of which clearly support the Deconcentration Perspective. However, across all regions non-metro counties grew faster than metro counties during this time period. Other studies went on to examine the late 1970s and found that "population deconcentration or suburbanization has not reversed, and there is no evidence of faster growth than in the suburbs" (Edmonston and Guterbock 1984). However, Frey ended his study by admitting that in the early 1980s these trends began to slow down and stated that "it remains to be seen whether or no these deconcentration tendencies will lead to continuing depopulation of the metropolis" (1987).

A modest array of studies have used geographies at the sub-county level, such as Johansen and Fuguitt (1984) and Luloff (1990). These studies have focused on the smallest size population settlements, villages, and, owing largely to the technical labor involved, included only a sample of all such settlements in the U.S. Luloff's (1990) study was focused on the changing number of small towns and larger places as well as their resident populations. His emphasis was on the linkage of place-based population change to the presence of natural resources and extractive industries.

A very few studies have included all incorporated places of 2,500 population and above and coupled them with the counties in which they are located (Lichter and

Fuguitt 1982; Fuguitt and Lichter 1989). These latter two efforts inform us of population dynamics within counties in non-metropolitan areas. Their foci largely point toward the population "deconcentration" hypothesis and do not include data beyond the 1984 population estimates (Fuguitt and Lichter 1989). However, the approach taken by Lichter and Fuguitt serves as a point of departure for our study of population dynamics in the rural-urban continuum.

Lichter and Fuguitt (1982) combine both non-metropolitan county and incorporated place-level population data, for incorporated places 2,500 and over, into a consolidated framework for analyzing change within counties. Their temporal coverage included 1950, 1960, and 1970 data from Census files with 1975 population estimates. Fuguitt and Lichter (1989) used 1960, 1970, 1980 Census data with 1984 population estimates. Using counties in the conterminous U.S. as their unit of analysis, these two studies configure county populations into two segments: (a) *urban population*, the sum of persons residing in all places of 2,500 or more; and (b) *rural population*, the remainder of the county population taken as a residual. Unincorporated places of 2,500 and above were not included. Annualized growth rates were calculated for the urban and rural population segments and a measure of "deconcentration" was computed by subtracting the urban rate from the rural rate. Note that this procedure is *aspatial* in nature. The procedures updating these results through 1984 are virtually identical. Lichter and Fuguitt (1982) reported that post-1970 trends showed a marked deconcentration within non-metropolitan counties and, based upon a regression model's results, that this pattern of deconcentration was increasingly less related to a set of traditional ecological, economic, and demographic variables. However, the clear findings in both studies was that non-metropolitan areas experienced marked patterns of population deconcentration. Fuguitt and Lichter (1989: 95) concluded, "It seems remarkable that in the 1970–1980 period more than one-half of the more rural nonadjacent counties experienced faster rural than urban growth." Their results for the early 1980s (through 1984) showed that some concentration was observed in counties with a city of 10,000 or more population. These results tended to be observed in all four regions of the U.S.

10.2.2 Defining Rural, Urban, and Community

Many scholars have debated a definition of "rural" America and this debate has largely involved a parallel concern with definitions of rural "communities" (Wilkinson 1991). Whitaker (1982) reports that the term rural was first used by the Bureau of the Census in 1874 with a definition of a *residence outside of cities or towns with 8,000 or more residents*. We emphasize this original definition by the Census Bureau for reasons we point out below. Ricketts et al. (1998) provide a comprehensive review of the various definitions for determining rural areas in the U.S. while Wilkinson's book (1991) grapples with similar variations for identifying social communities in rural areas. Using the "field theory" approach to community that he developed with Harold F. Kaufman, Wilkinson argues that to confuse the rural-urban continuum with a "past-present" continuum, a type of "cultural lag" domain assumption, has been part of the problem. The connection between "rural" and "community" is an intimate one:

> Rural. . .is a territorial concept. This is a most important consideration..because the community..has a territorial base. The study of rural life and community, therefore, is the study of the associations between one essential element of the community (i.e., the territorial element) and other essential elements of the community. The territorial concept of rural needs further specification and refinement to be useful in sociology. The land itself is not the point of sociological interest. What is of interest is the arrangement of people and activities on the land. Rural, as a sociological variable, refers to the extent of dispersion of people in a local ecology. Dispersion is of sociological importance because of its presumed effects on the interactions of people. (1991: 57).

Thus, according to this line of reasoning, we can expect that rural locales may be ecologically-definable but that they may not contain singular communities per se.

The various definitions of rural locales reviewed recently by Ricketts et al. (1998), for instance, show that governmental agencies and researchers define rural areas in widely divergent ways, begging the question of what is the phenomenon being classified in each rural-urban taxonomy. Attempts to extend the metropolitan vs. non-metropolitan dichotomy at the county-level are reflected in the long-standing work by Beale and his colleagues at USDA who effectively added non-metropolitan county adjacency to MSAs with the rural-urban continuum classification for counties (Butler and Beale 1994). This classification has been periodically updated and recently complemented by the urban influence taxonomy (Ghelfi and Parker 1997). The urban influence classification system essentially refines the Beale codes by segmenting the MSAs differently by size and adding a distinction among non-metropolitan counties of the presence of a city size of 10,000 persons or more. These two county-based taxonomies have facilitated a better understanding of population and other dynamics in non-metropolitan counties through a greater classification precision of a rural-urban continuum. Nonetheless, they are limited to the county-level and suffer from the varying geographic sizes of these administrative boundaries for many research purposes (e.g., Lichter 1993; Morrill et al. 1999).

While a number of sub-county geographies could be used in order to examine sub-county population dynamics, there are a number of problems with each of them. Census tracts and block groups have been used but they have a definite urban bias towards them in that their areal size gets larger as population density declines. Neither easily translate to proxies for rural communities. The US Postal Service's venerable zip code areas are both far too unstable in terms of their changes over time and the use of "point" zipcodes for large institutions, such as corporations, universities, and so forth. This study introduces a new sub-county geography (Non-Place Territory "NPT") for the purpose of identifying population dynamics along the urban/rural continuum. This geography is both easily understood and applicable, as it is consistent on census definitions and change over time.

10.2.3 Scale and the Place-Level Spatial Mobility of Population Data

The ecological dynamics of criminal offending and its spatio-temporal trends are directly impacted by the geographic scale of the area of interest (Agnew 1993). Mobility processes that help disseminate, or are directly concerned with the spatial mobility of a social issue or innovation, occur at many different geographic scales and can be quite different based on the resolution used in the study (Alber et al. 1992). However, as the modern world has become more urbanized, and made up of large consolidated aggregates of individuals, spatial mobility has taken on an "oozing" dynamic associated with the contagious spread of processes from one area to another (Alber et al. 1992). The globalized patterns brought to light by Wallerstein (1974, 1980, 1989), help to set the framework for place interactions at lower levels of geography. Furthermore, from this point of view, it is evident that places tend to perform some sort of function for one another, meaning that the relationship between them can be viewed as structural (Agnew 1993).

Often studies of U.S. crime are relegated to heterogeneous units of analysis, such as counties, or minute portions of the country, such as local examinations of tract and block data. This study introduces the examination of these spatio-temporal patterns at substantively meaningful sub-county geography at a national scale through the implementation of a place-level examination of reported crime. A half-century ago, Esselstyn (1953) called for the development of a "geographically non-urban" criminology. Esselstyn was primarily focused on the development of a conceptualized space, resulting in the development of the term "open country" used to describe any area not under some form of place-level police (and by inference, other city-based) jurisdiction. Since this early call for a better understanding of the geography of crime, which is included in the ecological analysis of crime, we must point out that there has been substantial discourse on the constitution of urban and rural, in relation to a number of demographically pertinent issues. Among these are how to include space into such analyses as well as the appropriate geography upon which to base these inquiries.

Here, it is made evident that the spatial mobility of criminal processes can be identified and examined at various spatial resolutions (Agnew 1993). Furthermore, each of these resolutions tends to illicit a somewhat different understanding and potential analytic problem of the process at hand, whether it is from dilution of variation and activity on a large scale or a misidentification of the process on small scales (Alber et al. 1992). Also, it is evident that the 'mobility' of social processes tends to be downward in the sense that core areas tend to send information and ideas to periphery areas (Wallerstein 1974, 1980, 1989; Agnew 1993). In the examination of crime, this downward/hierarchical 'spreading' process is most commonly concerned with the outward mobility of crime from a core central location to more periphery surrounding locations. Grounded in this framework, it is expected that criminological patterns spread between urban cities, along with innovation, ideas, etc., and the rural hinterlands.

10.2.4 Concept of Non-Place Territory

The approach taken in this study was developed by Howell (2004) and is consistent with both Wilkinson's (1991: 57) call for further "specification and refinement" of the territorial conceptualization of rural locales as well as Tickamyer's urging of attention to the measurement of space in rural studies. We examine rural-urban population dynamics during the 1980–2010 study period, but integrate both county and place-level data into a multi-level spatial framework. Using a decomposition of county population data into its constituent "place" and "non-place" parts, we operationalize a new territorial concept for rural locales, the *non-place territory*. We use geographic information systems (GIS) procedures to assist in the construction of the requisite data as well as to visualize some of the results. Through the multi-level linkage of place and county data, we examine trends in the growth rates of counties, their constituent places, and the segment of their population residing "non-places". The result, illustrated in Fig. 10.1, is an operational example of a territorially-based concept of diverse types of rural areas in the U.S. that can be implemented backward through several Decennial Censuses and forward through future ones (see Bureau of the Census 1994).

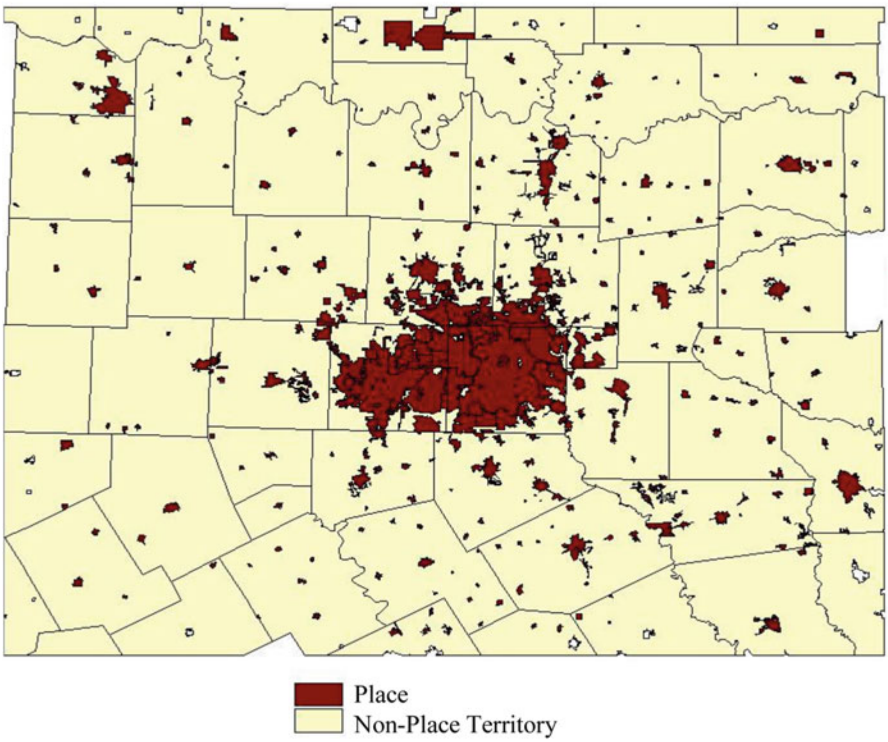

Fig. 10.1 Example of place/non-place geography, Dallas-Fort Worth metro area

Table 10.1 Number and dynamics of places[a] and non-place territory

	Study period:		
	1980–1990	1990–2000	2000–2010
Scope of study – IN – IN (area existed in both time periods)	25,048	26,507	27,556
Place deaths – IN – OUT (area did exist in the beginning of the time period but not in the end)	1,100	623	464
Place births – OUT – IN (area did not exist in the beginning of time period but did at the end)	2,559	2,726	1,513

[a]Figures representing places are actual change in place-parts

It is important to note that, as with other sub-national geographies, places dynamically change over time. At times places annex other non-place land or are established in areas where no place existed before. Also, there are cases where places cease to exist in physical space as they once did. This is due to a combination of phenomena, including being swallowed into larger places or, on the other end of the spectrum, being abandoned. Table 10.1 presents a description of the sample employed for the analysis that follows in this chapter. In each decade there are a number of place "births" and place "deaths". In order to conduct an examination of the spatial mobility of populations within counties, these "births" and "deaths" of places were restricted to the known places at the baseline for each of the three decadal analyses (80–90, 90–00, and 00–10).[1] The study units are presented in Table 10.1 with a count of the numbers of place "births" and "deaths". From this table one can see that the number of place-level units increased incrementally across the three study time periods with 25,048 place-level units in the 1980–1990 time period and 27,556 place-level units in the 2000–2010 time period. The table also indicates that place "deaths" have decreased over the 40 year period while place "births" in the 2000–2010 period were much lower than the observed number of "births" in the preceding study periods.

10.3 Research Methods

10.3.1 Source of Data and Variables

Population data for this study were obtained from the decennial census for 1980, 1990, 2000, and 2010. Since we are dealing with sub-county units of analysis, data were obtained at both the census place level (place parts for within county aggregation) and county level. Particularly, we are interested in the decennial growth

[1] The Geographic Areas Reference Manual from the U.S. Census Bureau says that these place "births & deaths" may take place for a number of reasons including consolidation, annexation, or detachment. In all, place births and deaths made up less than 3 % of the units.

rates of the individual unit's population change and, for standardization purposes, in the share of the encompassing county's total population. The idea of using NPT as a sub-county geography measure can be implemented easily with count variables, such as population counts and population changes. The formula is easily computed as the original geographies total count minus the sum of all sub-geographies. In the case of NPT territory at the county level, that would mean the county total for population minus the sum of the population of all places within that county.

$$\text{NPT} = \text{County Total} - \Sigma(\text{Place Totals})$$

From this formula, anything left over is not considered place population by the census; therefore through simple process of elimination, it is non-place population.

The use of spatial analysis with NPT, involves first the creation of Non-Place Territory GIS coverage. In order to create this coverage TIGER cartographic boundary files were obtained via the Census' web page. Files were obtained for 1980, 1990, 2000 and 2010, and they included the respective county and place-part files for each year. Place-parts were used in order to allow for the division of population in each place that crossed county lines. This allowed for the county specific counts of population.

Next, the GIS coverages were matched by year (i.e., 1980 county with 1980 place-parts) and the places were cut from the county coverage using a clipping technique. The resulting file is a complete county file with holes representing the area in which census defined places lye. This then is a comprehensive NPT geographic coverage as it represents the entire county that is not accounted for by Census defined places. In order to perform this this phase of analysis, these place-parts were then spatially merged back to the clipped NPT coverage, resulting in a seamless coverage of places and non-place territory with a FIPS-based record identification structure that included a five digit county FIPS for the NPT and a nine-digit place FIPS for the places.

Population data obtained via the above formula was then computed for each NPT and joined along with place population data to the merged place-level geographic coverage. Ultimately, this resulted in a geographic coverage of 23,435 geographic units with population count data for 1980, 1990, 2000, and 2010. Population data for the 23,425 units were then examined in both raw count and "share-of-county" form through the use of descriptive statistics and exploratory spatial data analysis (ESDA) procedures. Descriptively, these counts and proportions were examined across all time periods (including the entire study period) and across the place/NPT delineation within the place-level geography.

10.3.2 Measures

County and Place-level population count data from 1980, 1990, 2000, and 2010 were used to compute the two variables in the analysis. First, a measure of population change was computed for each of the decade time periods by subtracting

Table 10.2 Within and between county variance in population change by decade[a]

Time period	Between county variance (%)	Within county variance (%)
1980–1990	9.9	90.1
1990–2000	7.9	92.1
2000–2010	3.8	96.2

[a]Within and between county variance computed per Eta-Squared statistic from Univariate General Linear Model

time 1 (t_1) from time 2 (t_2) and dividing that result by the base of time 1 (t_1). This proportional change was converted into a percentage to be examined as an initial indicator of evidence for the utility of the place-level geography. Results from an Analysis of Variance are presented in Table 10.2. The table indicates that in all three decades the less than 10 % of the variance associated with county level population change can be accounted for at the county level (between counties). In fact, the trend seems to indicate that over time, the county has become an even less useful container of population counts with the 2000–2010 result indicating that only about 4 % of county level population change can be accounted for between counties. In contrast, greater than 90 % (96 % from 2000 to 2010) of county level population change can be attributed to *within-county* variation.

Furthermore, the results from Fig. 10.2 indicate that there are noticeable differences in population change across metropolitan status categories. From the bar chart, we can see the annualized percent change for places and NPTs that are in counties characterized as “metro”, “adjacent”, and “non-adjacent”. The 1980–1990 panel indicates that population counts in places and NPTs grew in metro and adjacent to metropolitan counties, but decreased in non-adjacent to metropolitan counties as an overall trend. The final grouping of bars represents the difference in growth across NPTs and places (place-NPT) with a negative value indicating growth at a faster rate in NPTs or loss at a slower rate in NPTs. It is clear that, on average, NPTs faired better on an annualized percent change basis then places in regards to growing their populations. We see a similar pattern from 1990 to 2000 where NPTs and Places grew across the board, but NPTs grew at a faster rate (again with negative difference values being presented). Finally, from 2000 to 2010 the pattern remains constant with places remaining basically stagnant, but all other categories growing in population. Again, the negative difference values indicate that NPTs grew at a faster annualized rate than did places.

The second variable computed for this analysis is an intra-county share ratio measure. To compute the variable, share ratios for all place-level units were divided by the total population of the containing county. Once these share where computed, the time two (t_2) share of the specific decade was divided by the time one (t_1) share. The resulting variable is a share ratio that indicates growth in the specific unit for all values above one, loss in the share of the county’s population for values below one, and no change for values at one (indicating t_2 and t_1 are equal).

As with the percent population change variable, one can see differences across both metropolitan status and place-level categories. The results indicate that in

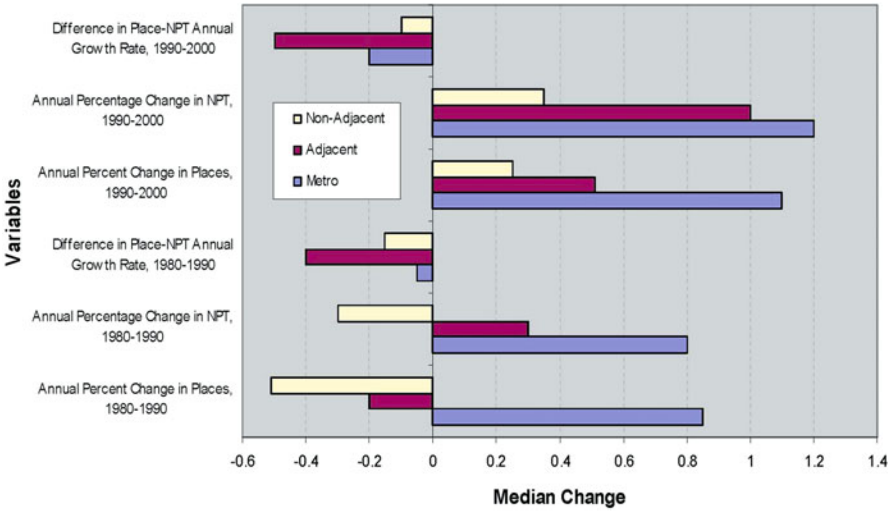

Fig. 10.2 Annualized percent change in place-level population by metro status, 1980–2010

1980 places in metropolitan counties accounted for about 1.3 % of the county population while NPTs accounted for about 35 %. In Adjacent counties places accounted for 2.2 % while NPTs accounted for about 51 % and in non-adjacent counties places accounted for about 3 % and NPTs for about 51 %. The share of intra-county population looks relatively consistent over time at each of the static points (1980, 1990, 2000, 2010). Towards the bottom of Table 10.3, we can also see that the patterns of change in the population share were also relatively consistent across decades with places generally losing shares of population and NPTs gaining in shares. This pattern has increased in metropolitan counties in the most recent decade while patterns in adjacent and non-adjacent counties have changed little over the time period. Overall, it seems that the largest discrepancies have traditionally been found in adjacent to metro counties, with the most recent period (2000–2010) finding metropolitan counties out-pacing adjacent counties in NPT growth for the first time (Table 10.3).

10.3.3 Identifying Within-County Population Centers

The results derived from the statistical techniques employed in this analysis are spatial and as such are sensitive to the definition of the neighborhood and the resolution at which the social process of interest is examined (Anselin 1995; Agnew 1993). This makes it important to understand the potential connectivity of places prior to the implementation of the analysis. Furthermore, it is important to define your given neighborhood as being grounded in some theoretical framework, which in this case is interested in the identified relationships of articulation between neighboring places

Table 10.3 Median proportion of population (Share of Total County) by place status *within* county and change in proportion *within* county by decade, 1980–2010[a]

	Metro		Adjacent to metro		Non-adjacent to metro	
	Place	NPT	Place	NPT	Place	NPT
Share of population						
1980 share	.013	.345	.023	.513	.032	.508
1990 share	.013	.342	.022	.508	.030	.513
2000 share	.014	.357	.022	.532	.032	.517
2010 share	.012	.368	.020	.533	.030	.528
Share change						
1980–1990	−.020	.000	−.060	.013	−.039	.009
1990–2000	−.041	.004	−.063	.007	−.035	.011
2000–2010	−.071	.014	−.054	.003	−.034	.006

[a]Each County includes 1 NPT (balance of county) and multiple places

and non-places within the same county (Waller and Gotway 2004). In ancillary analyses, the spatial population centers (perhaps serving as ecological "neighborhoods") were defined using a number of differing approaches in order to maximize the within-county relationships (Anselin 1995). Maximizing the within county connectivity is important due to the fact that one of the goals of this paper is to identify patterns of concentration and deconcentration within the same county. Implementing some of the work outlined above, the transmission of social processes, behaviors, and information is often found to take place in a core-to-periphery fashion (Agnew 1993). It is evident that the mobility of population should act in the same fashion (the goal then is to capture the average, most common, connectivity definition between places understanding that it is not always the case).

For this purpose a *k*-nearest neighbors approach was employed.[2] By aggregating (summing) the number of places within a given county and computing simple descriptive statistics on that count, it is possible to identify potential *k's* to be used in the definition of the within-county neighborhoods. The range of places within a county varies greatly from zero to seventy-seven, with a mean of 2.75, a median of 2.34, and a standard deviation of 3.01. However, they concentrate heavily between two and four, with 70 % of the counties having two places, 81 % having two or three places, and 87 % having between two and four places.

Ultimately, four was chosen as the number of nearest neighbors for each locality.[3] Since the non-place is the focal point of the change, this nearest neighbors approach looks to be more efficient in identifying within-county neighborhoods when compared to the "queen's matrix" because, on average, the places within the

[2] The *k* nearest neighbors approach identifies a theoretically grounded number of meaningful neighbors based on locality centroids and Euclidean distance (Anselin 1995).

[3] For sensitivity purposes, the analyses were run with k = 2, 3, and 4. k = 4 was ultimately chosen based on the balance between meaningful significant results compared to k = 2 and k = 4. Distance based and contiguity based matricies were tested as well with the k-nearest neighbors approach ultimately proving to be the most theoretically and empirically appropriate definition tool.

county will be the non-places only neighbors. There will be some instances where there are less than four places, in these cases a place from neighboring non-place, or the neighboring non-place itself, will be included as a neighbor. This will lead to a few instances where between-county distribution may be identified. However, from the simple statistics above, one can see that this will be the exception as opposed to the norm. For the purposes of maximizing the within-county connectivity and following the results of ancillary analyses, it seems that the $k = 4$ nearest neighbors approach is the most efficient definition.

10.3.4 Analysis

In analytical terms, this study builds on the previous applications of the univariate LISA statistic as an efficient identifier of statistically significant patterns of spatial mobility associated with population concentration and deconcentration in the US over the past 40 years. Building upon the innovative work by Cohen and Tita (1998) and Porter (2010, 2011), this analysis examines intra-county population share changes over a decade (via the share-ratio variable outlined above) using a univariate examination of spatial autocorrelation via the Moran's I statistic:

$$I = \left(\frac{1}{s2}\right) \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \omega ij\left(Xi - \overline{X}\right)\left(Xj - \overline{X}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \omega ij}$$

where:

$$s2 = \frac{1}{N}\sum_{i=1}^{N} \left(Xi - \overline{X}\right) \wedge 2$$

This equation compares the share ratio of any given place-level unit (local unit i) to the average share ratio of all four neighboring units (neighborhood unit's j). The results of the univariate Moran's I test on the share ratio (measuring change in share over the 10 year period) are presented in Table 10.4. The results indicate that there are significant patterns of spatial connectivity associated with the share ratio and the 4 nearest neighbors. Given the conceptualization of the definition as a measure of intra-county connectivity and the significant negative values of the Moran's coefficients, the global trend seems to be one where as the share ratio increases in the local area there is a loss of share in neighboring areas. Also, as the share decreases in a local area there is a gain in neighboring areas. This is generally referred to as

Table 10.4 Univariate Moran's I or change in population share by decade, 1980–2010

	1980–1990	1990–2000	2000–2010
Moran's I	−0.0873***	−.1760***	−0.1296***

***p-value < 0.001

negative spatial autocorrelation. It seems to support the theoretical concept of concentrating and deconcentrating populations (gaining or losing population in a zero-sum form related to neighboring units).

Next, local areas of significant clusters were identified via the univariate LISA statistic in order to measure the proportional contribution to the overall levels of spatial dependence in the data with a significance value (Waller and Gotway 2004). The equation for the univariate LISA as it is employed is as follows:

$$Ii = \sum_{j=1}^{N} \omega ij\left(Xi - \overline{X}\right)\left(Xj - \overline{X}\right)$$

From the equation, one can see that the random variable, I_i, is equal to the spatial weight indicator (neighborhood connectivity matrix), multiplied by the product of the share ratio of the local unit (X_i) and the average share ratio of the neighborhood (X_j), summed across all neighborhood units (j). This approach will, then, allow for the examination of pockets of significant spatial change in regards to intra-county population. Ultimately, the LISA analysis results in a series of five possible outcomes. First the local unit may not be in a significant spatial cluster. Second, the local unit may have a high share ratio and the neighborhood average may also be report a high share ratio (High-High). Third, the local unit may have a low share ratio and the neighborhood average may also be low (Low-Low). Fourth, the local unit may have a low share ratio but the neighborhood average share ratio may be high (Low-High). Fifth, the local unit may have a high share ratio but the neighborhood average share ratio may be low (High-Low).

10.4 Results

The results of the LISA analysis required a differentiation across place and NPT categorizations. For instance, an NPT that reports a Low-High LISA result lost population while the neighborhood (local places) gained in share ratio. This would indicate growth in places and loss in NPTs or concentration. However, a place that lost while the units around it gained (Low-High) would likely be deconcetrating. Universal growth (High-High) and universal loss (Low-Low) do not need to be differentiated by place level.

These LISA results relate to population distribution in which place growth outpaces NPT growth (a concentrating population trend) or NPT growth outpaces place population growth (a deconcentrating population trend). Illustrations of both are presented in Figs. 10.3 and 10.4. Figure 10.3 is of Guilford County, NC which includes the place of Greensboro. The map illustrates population concentration as Greensboro increased its population by 4.48 % while the larger county of Guilford lost population at a rate 6.23 %. Figure 10.4 represents the opposite phenomena where the county, Knox County, TN in this case, gains population while the places within the county lose population. From Fig. 10.4 one can see that while Knox County, TN's population increased by 5.51 % from 1990 to 2000, the city of Knoxville lost population at a rate of 6.03 %. *The term deconcentration then simply refers to the phenomena where NPTs gain and places lose, while the term concentration simply refers to the phenomena where NPTs lose and places gain.*

The LISA results are presented in Table 10.5. The first set of rows refers to places and NPTs in metropolitan counties across the three time periods. The results indicate universal growth (High-High) in Metros, more often in places than NPTs in all three time periods, but that that rate of growth slowed down from 1990 to 2000 and 2000 to 2010. In contrast, universal population loss in metro counties was lowest from 1980 to 1990 and picked up over the latter two decades. The Low-High row must be examined conversely as concentration for NPTs and deconcentration for places. The results are relatively consistent across time periods, but suggest that the rate of deconcentration for places in metro counties increased in the latter two decades. The High-Low results must be differentiated as deconcentration for NPTs and concentration for places. Likewise, we observe deconcentration increase in the latter two decades for NPTs in metropolitan counties. In metropolitan counties, the dominant form of significant spatial population redistribution is one of deconcentration across all three time periods.

In adjacent to metropolitan counties, the dominant forms of population redistribution vary by decade. From 1980 to 1990, deconcentration and high rates of universal loss are observed. However, from 1990 to 2000 and 2000 to 2010, concentration became much more prevalent. In non-adjacent counties, the dominant form of population redistribution came in the form of concentration in from 1980 to 1990 but then deconcentration from 1990 to 2000 and 2000 to 2010. This indicates that adjacent and non-adjacent counties operated opposite of each other in terms of the timing of their concentration and deconcentration trends.

10.5 Discussion

We offered a new geography in this chapter, that of the *non-place territory*, to identify sub-county land-mass commonly called "out in the county" by residents. The call for this volume asked for contributions that might further the cause toward measuring the Holy Grail of social communities. We clearly have not accomplished the goal at this point but have offered a reasoned operational definition of a new sub-county geography that holds meaning for residents. In comparison to the use of Census tracts, for instance, can the reader report the *exact* tract in which s/he lives?

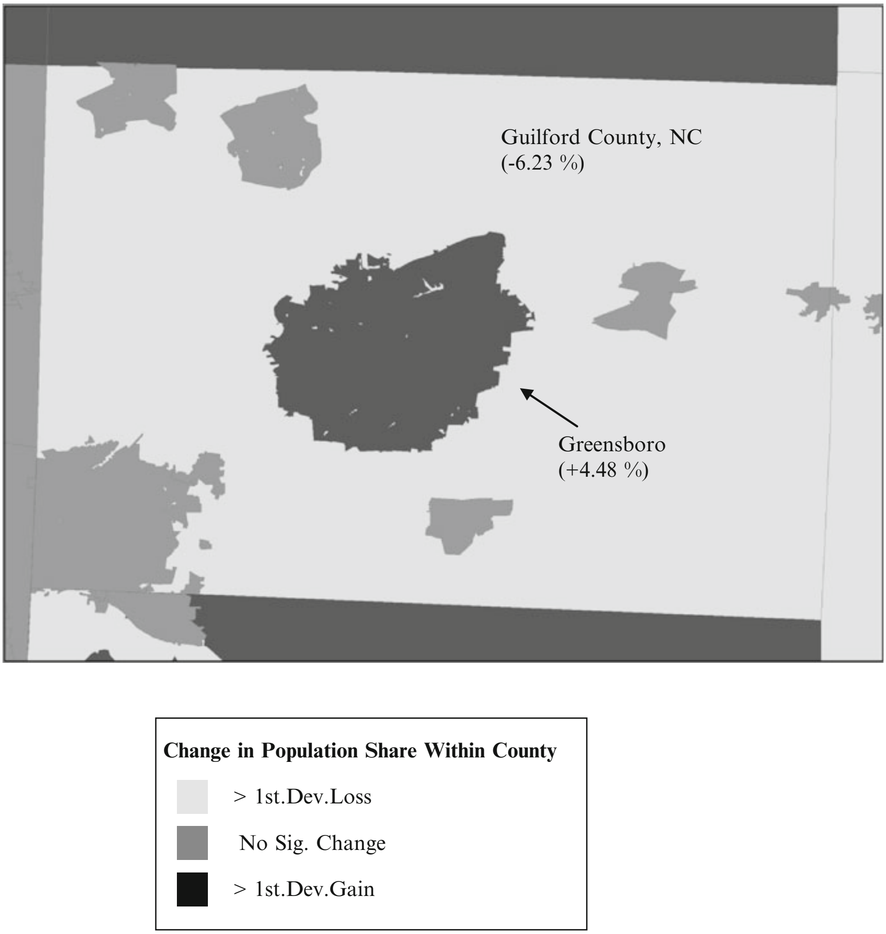

Fig. 10.3 Example of significant instance of within-county population concentration as measured via change in share of county population; Greensboro, NC (1980–1990)

But, by comparison, does *any* reader not know whether s/he lives in an incorporated place or out in the county? This operational definition can be replicated in all Census geographies since 1980 as that was the first decennial census in which places were spatially digitized and released to the public.

Our empirical results track the patterns of population concentration in what are already urbanized localities. They also identify the intriguing pattern of population deconcentration into what are (or were) typified as rural localities. In place of judging percentages of population change, we added a significance test using replication in Anselin's LISA statistic. This adds greater objectivity to the characterization of population redistribution patterns at the county level and smaller areas. It also passes what we refer to as the *umbrella test* in spatial demography. Like a statewide weather report, the conventional demographic result of metropolitan or

Fig. 10.4 Example of significant instance of within-county population deconcentration as measured via change in share of county population; Knoxville, TN (1990–2000)

Table 10.5 LISA clusters *within* place-level by metropolitan status, 1980–2010

	% LISA cluster *within* place-level classification					
LISA clusters	1980–1990		1990–2000		2000–2010	
Metro	NPT	Place	NPT	Place	NPT	Place
High-High	2.1[a]	5.9[a]	1.7[a]	2.8[a]	1.7[a]	2.8[a]
Low-Low	4.6[b]	7.4[b]	6.6[b]	12.7[b]	6.6[b]	12.6[b]
Low-High	7.2[c]	3.8[d]	7.4[c]	5.5[d]	7.4[c]	5.5[d]
High-Low	6.7[d]	2.5[c]	8.6[d]	1.0[c]	8.6[d]	1.0[c]
Adjacent to metro						
High-High	0.0[a]	3.1[a]	2.2[a]	4.8[a]	2.2[a]	4.2[a]
Low-Low	8.9[b]	8.3[b]	1.1[b]	6.2[b]	1.1[b]	6.4[b]
Low-High	4.4[c]	2.5[d]	10.0[c]	4.2[d]	10.0[c]	4.2[d]
High-Low	10.0[d]	2.3[c]	5.6[d]	1.0[c]	5.6[d]	1.1[c]
Non-adjacent						
High-High	1.1[a]	0.8[a]	0.3[a]	0.7[a]	0.3[a]	0.7[a]
Low-Low	1.2[b]	2.6[b]	1.1[b]	2.0[b]	1.1[b]	1.9[b]
Low-High	1.6[c]	1.1[d]	3.7[c]	2.2[d]	3.7[c]	2.2[d]
High-Low	4.1[d]	1.5[c]	2.6[d]	0.4[c]	2.6[d]	0.3[c]

[a]Regional growth
[b]Regional loss
[c]Concentration
[d]Deconcentration

rural growth and decline tells us little about "where" the gains or losses occur within the much larger landscape of a state. Thus, a smaller scale to seek changes within a county comes closer to population change in a locality's "neck of the woods". This issue begs the question of the conventional modifiable areal unit problem (MAUP; see Gehlke and Biehl 1934; Openshaw 1983). Criticisms of the county scale partly led to our offering of the place vs. non-place *territory* geography in this chapter.

While these results do build upon those of previous demographers studying domestic population redistribution, we believe that adding the finer point of spatial proximity and specificity of county and *within*-county change moves prior work into a more effective framework for future studies. Indeed, the simple ANOVA results revealing the much greater variation in population change within counties than between counties in all three decades underscores the utility of this initial approach.

This sub-county dissection requires future work to further characterize the *non-place territory* in the search for localities which may be "social communities". This approach does not vacate work to be done *within* places using tracts, block-groups, or blocks. Much like Porter and Howell (2009) documenting the heterogeneity of metropolitan areas where agricultural production typifying "rural" localities occurred substantially within the boundaries of official metropolitan areas (see also Thomas and Howell (2003), we believe that the "community field" approach developed by Kaufman and his student Wilkinson (Wilkinson 1991), essentially combining Galpin's early institutional mapping approach with the psychologist Levin's spatial cognition theory, can be applied to the further characterization of sub-place and non-place territory. This should be an explicit goal of future studies of this genre.

Some years ago, Isserman (2001) showed the encroachment of metropolitanization into non-metropolitan areas was through the previously non-metro counties subsequently becoming classified as metropolitan by the OMB process (see also Bureau of the Census 1994). The *mechanisms* through which this "metropolitanization" occurs (e.g., commuting flows from growing rural residential enclaves) has yet to be spatially described for the U.S. We call these mechanisms *spatially-expressed social bonds* owing to Galpin's social institution participation boundaries in Walworth County, WI (Galpin 1915) and Robert E. Park's mapping of newspaper subscription areas for the greater Chicago area (Park 1929). We believe that further articulating and identifying these spatially-expressed social bonds at the most effective spatial scale will further our search for the Holy Grail of an ecological, non-transient operationalization of something we've called community for a century now. The challenge of multiple spatial scales, long realized by rural scholars (e.g., Greenwood and Luloff 1979), and the juxtaposition of "bottom-up" measures of interaction vs. "top down" ecological measures still face us.

Future work underway involves amplifying the Field Theory approach initially described by Harold F. Kaufman (1959) and his principal student, Kenneth P. Wilkinson, but continued by Alvin E. Luloff (e.g., Luloff and Wilkinson 1977; Bridger and Luloff 1999) using Wilkinson's rubric, positively affirm a locality as rural (Wilkinson 1991). Wilkinson argued that the Census Bureau definition of typifying land mass as rural as that not positively attributed as urban was misleading. He articulated that overlaying the population distribution over how land is utilized was the key element for positively attributing rurality to a land mass.

This work must cross a great intellectual divide between the mainstream demography literature and scholars of rural settings (e.g., rural sociology). We have consistently observed that urban demographers simply do not read or cite the voluminous literature available to them by their rural counterparts (see Porter and Howell 2009). The origins of the Chicago School's Social Ecology scholarship was clearly the earlier work by Charles J. Galpin (see Park 1929) who used maps to spatially describe the "attendance" or "service" zones of key social institutions in a rural Wisconsin county (Walworth WI). Much of the divide between those scholars who argue that "community" should be built from the "ground up" (e.g., Grannis 2009), often through volunteered or "crowd-sourced" data (e.g., Sui et al. 2012), and those conventional demographic analysts using Census Bureau-centric measurements is bridged through the long-standing work by rural community and associated scholars who have endeavored to understand social "attachment," "satisfaction," "identification" and other subjective elements of localities. The importance of friendship networks (Freudenburg 1986), for instance, was made clear decades ago and perhaps crowd-sourced data would yield measurements or proxies of such networks. Even the well-known Field Theory approach to community pioneered in the 1950s and 1960s by Kaufman and his student Wilkinson in Mississippi went unmentioned by the most positive assessment of Field Theory by Martin (Martin 2003). Until an intellectual bridge between mainstream and rural scholars is built – and we believe that the bias is from mainstream toward the rural – these advances will not likely take place.

References

Agnew, J. (1993). Representing space: Space, scale and culture in social science. In J. Duncan & D. Ley (Eds.), *Place/culture/representation* (pp. 251–271). London: Routledge.

Alber, I., Bassani, J. L., Khantha, M., Vitek, V., & Wang, G. J. (1992). Grain boundaries as heterogeneous systems: Atomic and continuum elastic properties. *Philosophical Transactions of the Royal Society A, 339*, 552–586.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.

Bridger, J. C., & Luloff, A. E. (1999). Toward an interactional approach to sustainable community development. *Journal of Rural Studies, 15*(4), 377–387.

Brown, D. L., & Zuiches, J. J. (1993). Rural-urban population redistribution in the United States at the end of the twentieth century. In D. L. Brown, D. Field, & J. J. Zuiches (Eds.), *The demography of rural life* (pp. 1–18). University Park: Northeast Regional Center for Rural Development.

Brown, D. L., & Wardell, J. M. (1980). *New directions in urban–rural migration. The population turnaround in rural America*. New York: Academic Press.

Bureau of the Census. (1994). *Geographic areas reference manual*. Washington, DC: Economics and Statistics Administration, Bureau of the Census.

Butler, M. A., & Beale, C. L. (1994). *Rural–urban Continuum codes for metro and nonmetro counties*, 1993. USDAERS; Washington, DC: 1994. AGES-9425.

Cohen, J., & Tita, G. (1998). *The gang-drug-gun nexus of homicide in Pittsburgh*. Working paper. H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA.

Dahmann, D. C., & Fitzsimmons, J. D. (Eds.). (1995). *Metropolitan and nonmetropolitan areas: New approaches to geographical definition* (Working Paper 12). Washington, DC: Population Division, Bureau of the Census.

Edmonston, B., & Guterbock, T. M. (1984). Is suburbanization slowing down? Recent trends in population deconcentration in U.S. metropolitan areas. *Social Forces, 62*(4), 905–925.

Esseltyn, T. C. (1953). The social role of a county sheriff. *Journal of Criminal Law and Criminology, 44*(2), 177–184.

Federal Register. (1999, October 20). Recommendations from the metropolitan areaStandards review committee to the office of management and budget concerning changes to the standards for defining metropolitan areas; Notice. P. 56636.

Freudenburg, W. R. (1986). The density of acquaintanceship: An overlooked variable in community research. *American Journal of Sociology, 92*, 27–63.

Frey, W. H. (1992). Metropolitan redistribution of the U.S. elderly: 1960–70, 1970–80, 1980–90. In A. Rogers (Ed.), *Elderly migration and population redistribution: A comparative perspective* (pp. 123–142). London: Belhaven.

Frey, W. H. (1993, April). US elderly population becoming more concentrated. *Population Today*, 6–9.

Frey, W. H. (1987). Migration and depopulation of the metropolis: Regional restructuring or rural renaissance? *American Sociological Review, 52*(2), 240–257.

Frey, W. H., & Spear, A., Jr. (1992). The revival of the metropolitan population growth in the United States: An assessment of findings from the 1990 census. *Population and Development Review, 18*(1), 129–146.

Fuguitt, G. V., & Lichter, D. T. (1989). Chapter 3: Small town growth and population dispersal. In G. V. Fuguitt, D. L. Brown, & C. L. Beale (Eds.), *Rural and small town America*. New York: Russell Sage.

Galpin, C. J. (1915). *The social anatomy of an agricultural community*. Madison: University of Wisconsin Agricultural Experiment Station Bulletin 34.

Gehlke, C. E., Biehl, K. (1934, March). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association, 29*(185A), 169–170.

Ghelfi, L. M., & Parker, T. S. (1997). *A county-level measure of urban influence* (Staff Paper No. 9702). Washington, DC: U.S. Dept. of Agriculture.

Grannis, R. (2009). *From the ground up: Translating geography into community through neighbor networks*. Princeton: Princeton University Press.

Greenwood, P. H., & Luloff, A. E. (1979). Inadvertent social theory: Aggregation and its effect on community research. *Journal of the Northeastern Agricultural Economics Council, 8*(1), 44–47.

Howell, F. M. (2004). *Spatial analysis in rural sociology*. Presentation at the 2004 meeting of the rural sociological society, Sacramento, CA.

Isserman, A. M. (2001). Competitive advantages of rural America in the next century. *International Regional Science Review, 24*(1), 38–58.

Johansen, H. E., & Fuguitt, G. V. (1984). The changing rural village. *Rural Development Perspectives, 6*, 2–6.

Johnson, K. M. (1989). Recent population redistribution trends in nonmetropolitan America. *Rural Sociology, 54*(3), 301–326.

Johnson, K. M., & Beale, C. L. (1994). The recent revival of widespread population growth in nonmetropolitan areas of the United States. *Rural Sociology, 4*, 655–667.

Kasarda, J. D., & Irwin, M. D. (1991). National business cycles and community competition for jobs. *Social Forces, 69*, 733–761.

Kaufman, H. F. (1959). Toward and interactional conception of community. *Social Forces, 38*, 8–17.
Lichter, D. T. (1992). Migration, population redistribution, and the new spatial inequality. In D. L. Brown, J. J. Zuiches, & D. R. Field (Eds.), *The demography of rural life* (pp. 19–46). University Park: Northeast Regional Center for Rural Development.
Lichter, D. T. (1993). Demographic aspects of the changing rural labor force. In L. L. Swanson & D. L. Brown (Eds.), *Population change and the future of rural America* (ERS staff report AGES, pp. 136–150). Washington, DC: Economic Research Service, U.S. Department of Agriculture.
Lichter, D. T., & Fuguitt, G. V. (1982). The transition of nonmetropolitan population deconcentration. *Demography, 19*(2), 211–221.
Luloff, A. E. (1990). Chapter 3: A social history of the small and rural community literature. In R. S. Krannich (Ed.), *The Lowry Nelson symposium: Rural villages in the twenty-first century*. Logan: Mountain West Center for Regional Studies.
Luloff, A. E., & Wilkinson, K. P. (1977). Is community alive and well in the inner-city? A comment on Hunter's loss of community. *American Sociological Review, 42*(5), 827–828.
Martin, J. L. (2003). What is field theory? *American Journal of Sociology, 109*(1), 1–49.
Morrill, R., Cromartie, J., & Hart, G. (1999). Metropolitan, urban, and rural commuting areas: Toward a better depiction of the United States settlement system. *Urban Geography, 20*, 727–748.
Openshaw, S. (1983). *The modifiable areal unit problem*. Norwick: Geo Books.
Park, R. E. (1929). Urbanization as measured by newspaper circulation. *American Journal of Sociology, 35*(1), 60–79.
Porter, J. R. (2010). *Tracking the mobility of crime: New methodologies and geographies in modeling the diffusion of offending*. Newcastle upon Tyne: Cambridge Scholars Publishing.
Porter, J. R. (2011). Identifying within-county spatio-temporal patterns of the articulated mobility of criminal offending: An application of multivariate spatial clustering techniques. *Systems Research and Behavioral Science, 28*(3), 197–211.
Porter, J. R., & Howell, F. M. (2009). On the 'Urbanness' of metro areas: Testing the homogeneity hypothesis, 1970–2000. *Population Research and Policy Review, 28*(5), 589–613.
Ricketts, T. C., Johnson-Webb, K. D., & Taylor, P. (1998). *Definitions of rural: A handbook for health policy makers and researchers*. Federal Office of Rural Health Policy, Health Resources and Services Administration, U.S. Department of Health and Social Services.
Sui, D. Z., Elwood, S., & Goodchild, M. F. (Eds.). (2012). *Crowdsourcing geographic knowledge: Volunteered geographic information in theory and practice*. Berlin: Springer.
Thomas, J. K., & Howell, F. M. (2003). Metropolitan proximity and U.S. Agricultural productivity, 1978–1997. *Rural Sociology, 68*(3), 366–386.
Vining, D., & Strauss, A. (1977). A demonstration that the current deconcentration of population in the U.S. is a clean break with the past. *Environment and Planning, 9*, 751–758.
Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data*. Hoboken: Wiley.
Wallerstien, I. (1974). Dependence in an interdependent world: The limited possibilities of transformation within the capitalist world-economy. *African Studies Review, 17*(1), 1–26.
Wallerstien, I. (1980). Development: Theories, research designs and empirical measures. In L. Blussé (Ed.), *History and underdevelopment* (pp. 21–28). Leiden: Centre for the Study of European Expansion.
Wallerstien, I. (1989). The ideological tensions of capitalism: Universalism versus racism and sexism. In J. Smith (Ed.), *Racism, sexism, and the world-system* (pp. 3–9). New York: Greenwood Press.
Whitaker, W. H. (1982). The many faces of Ephraim: In search of a functional typology of rural areas. 1982. ED 242 459.
Wilkinson, K. P. (1991). The rural-urban variable in community research. In *The community in rural America* (pp. 37–59). Middleton: Social Ecology Press.

Chapter 11
Socio-spatial Holes in the Advocacy Umbrella: The Spatial Diffusion of Risk and Network Response Among Environmental Organizations in the Marcellus Hydro-fracturing Region

Michael D. Irwin and Erin C. Pischke

11.1 Overview

This chapter examines the effect of the spatial distribution of hydro-fracturing activity in Pennsylvania on the formation of new coalitions among environmental advocacy organizations. These advocacy coalitions help mobilize environmental advocacy organizations across Pennsylvania. From a policy view, these environmental networks play a vital advocacy role, providing resources and information for local citizenry in the small towns and rural areas of Pennsylvania. In areas where these coalitions create effective organizational networks, advocacy resources are rich. Where such network connections are sparse, organizational resources are also sparse. The scope, structure and density of these networks constitute important dimensions of social movement studies and have been shown to be critical in movement success (Diani 2003a; McAdam and Boudet 2012; Gould 1991, 2003). The character and the formation of networks have emerged as a central variable in contemporary social movement literature (Diani 2003a, pp 10–11). Most of this literature has focused specifically on linkages among individuals and groups as mechanisms for the diffusion of ideology and strategy. Few have explored the relationship of these networks to the communities in which collective action takes place. McAdam's work is an exception (McAdam and Boudet 2012; McAdam and Fligstein 2012). As McAdam and Boudet state:

> Although useful, the work on movement diffusion suffers from two problems. First, as previously noted, the work is rarely conceptualized as research on the geographic expansion of a movement. Second, besides noting the structural conditions (e.g., preexisting

M.D. Irwin (✉)
Duquesne University, Pittsburgh, PA, USA
e-mail: irwinm@duq.edu

E.C. Pischke
Michigan Technological University, Houghton, MI, USA
e-mail: epischke@alumni.cmu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_11

> social ties) that facilitate diffusion, there is otherwise very little attention paid to the social processes that might play a role in movement expansion. (2012, p. 137)

Here the authors recognize the need for integrating network measures and for spatial concepts to understand the processes of collective action. In this chapter we will apply some of the most widespread models of socio-spatial interaction and network models to bear on the rise of a collective action network among communities. We will highlight some of the geographic elements of network formation and model the structure of these interactions. The role of distance on the spatial distribution of organizational interaction will be evaluated as well as the role that environmental impacts and demographic distributions play in creating this spatial pattern. We will begin by evaluating the spatial patterns of at-risk populations (demographic assessment). Then, our chapter will evaluate the implications of network formation for affected areas using measures of network centrality. Finally we will assess the role that distance, organizational distribution and the spatial pattern of fracking activity plays in this network structure.

11.2 Environmental Impacts and Organizational Responses

Early in the twenty-first century, technological innovations in natural gas drilling coupled with increased profits have fueled an explosion in exploration and drilling across the United States. This new hydraulic fracturing approach involves fracturing deep shale basins by drilling vertically to the depth of the shale, horizontally across the shale formation, then using specially formulated chemicals mixed with water to fracture the rock and release natural gas (Pennsylvania Department of Environmental Protection [DEP] 2012, pp. 2–3). The advantage of using horizontal fracking techniques over more common vertical wells is due to the capability of drilling out in several directions from one well as opposed to one direction of vertical wells; well pads are also constructed so that between two and ten wells can be drilled from each pad (DEP 2012; US Department of Energy [DOE] 2010, p. 4). Also, natural gas "exists in horizontal planes, [so] horizontal drilling increases the amount of penetration into the reservoirs" (Reeder 2010, p. 5). These gas deposits can be found in "shale basins" across the lower 48 United States (Kargbo et al. 2010, p. 5679). In Pennsylvania, a large swath of the Marcellus shale basin covers about half the state, roughly running in a crescent covering counties from the southwest of the state through the northeast of Pennsylvania (DEP 2010a, p. 1; Jacquet 2009; Kerr 2010, p. 1624). In the southwest of the state, Washington County, Pennsylvania, was home to the first Marcellus well in 2003, which began commercially producing gas in 2005 (Brasier et al. 2011, p. 33). The extraction of shale gas began in earnest in 2008 and has since steadily risen (Kargbo, et al. 2010, p. 5679).

From 2008 through 2009 the primary push from the gas industries was to purchase leases on private land in advance of getting drilling permits. Companies, along with industry associations, launched widespread community meetings to discuss the advantages to individuals and communities of leasing local land for fracking, often offering local grants to communities to help them prepare for the upcoming economic boom (Chief Oil and Gas 2009; *The Express* 2009). Rural communities and governments, as well as private land holders, celebrated this boom as an economic savior in an otherwise declining rural economy. Early estimates were that Marcellus drilling would add 29,000 jobs and $240 million in state and local tax revenue in 2008, with a cumulative addition of 175,000 jobs and $13.5 billion taxes over the next 10 years (Considine et al. 2009). Individual landowners were making windfall profits through leases. Local communities were writing agreements for local infrastructure development (mostly road repair) beyond their local budgets (Jacobson and Kelsy 2011; Murray and Ooms 2008, p. iv). The state government supported these efforts virtually without reservations, seeing this new industrial activity as a way to balance its budget through additional severance taxes on gas extraction. Initial reactions in the environmental community were somewhat ambiguous and organized opposition to fracking activity was not widespread. Natural gas production is not without an environmental upside. Increased availability of natural gas could provide a ready source of energy that might supplant more harmful energy sources including coal and oil (Cathles 2010, pp. 1–2; Soeder and Kappel 2009, p. 5). Potentially, shale gas could replace more environmentally impactful fuels and serve as a temporary bridge fuel to other alternative energy sources (MIT 2011). For some environmental organizations this possible role in a long-term energy solution moderated other environmental concerns (Kargbo et al. 2010, p. 5679; Kerr 2010, p. 1624).

In Pennsylvania the gas industry activity—leasing, permitting and drilling—really took off beginning in 2009. Prior to this time, these companies sought relatively few opportunities to lease land, then drill. In 2005 there were less than 20 Marcellus hydro-fracturing drilling permits sought. In 2008 this had risen to slightly under 600 fracking permits for drilling. By 2009 this had more than tripled (nearly 2000 permits), and in 2010 gas companies applied for more than 3600 drilling permits. By 2011 this had risen to nearly 4500 new hydro-fracturing drilling permits, covering virtually all Pennsylvania counties in the Marcellus Basin (DEP 2000–2012). Wherever permits were sought, landholders and communities were intensely engaged by the drilling companies in a flurry of lease purchases and local agreements with governments.

As gas companies ramped up the pace of land leasing and drilling across the Marcellus, various constituent groups within the State began to point out that such potential economic benefits could be counterbalanced by social and environmental burdens associated with natural gas drilling. Early on it became clear that few of the drilling jobs generated were held by Pennsylvanians (Fisher 2010). Few of the drilling firms were local. Most were companies shifting from the Texas or western shale basins (DEP 2010b). As a result, much of the economic impact flowed out of state. A number of sources raised doubts about the projected short term economic benefits for local communities (Christopherson 2011; Freudenburg and Wilson 2002; Headwaters Economics 2009).

Further, as fracking activity (both leasing/permits and drilling) geared up in 2009 and 2010, many citizen groups became more concerned about the environmental impacts. A few new groups arose, although these tended to be loose coalitions of opposition organizing for direct action protest (cf. Marcellus Shale Protest). More typically, existing groups, especially those with interests and concerns involving specific local watersheds, shifted ongoing activities to engage Marcellus issues. The environmental well-being of watersheds were a natural issue that encompassed a broad spectrum of civic organizations. The natural gas drilling process requires substantial use of local waters (up to seven million gallons per well) that these groups argued could result in severe impacts on both surface water and water tables. Drilling firms regularly drain local waters (streams, ponds, rivers) for use in the fracking process (DOE 2009). Additionally, the fracturing process requires that this water be infused with lubricants, biocides, rust inhibitors, solvents, etc. Most of this fracking 'slickwater' mixture is recovered from the well and stored on-site for reuse in plastic-lined pits. These fracking ponds can pose environmental and health risks to local populations and hold the potential for impacting adjacent streams and rivers (The Pittsburgh Geological Society n.d.; Schmidt 2013). These fresh water impacts are seldom contained in the areas proximate to the drilling sites, instead following the flow of fresh water throughout a region.

Disposal of the slickwater is eventually handled through evaporation, through injection into unused wells or mines or through disposal in local municipal wastewater treatment plants. As drilling activity increased, Pennsylvanian environmental groups began to catalog the many possible impacts this activity could have on local watersheds. They showed that many local water treatment plants were not equipped to treat fracking water, and these slickwater compounds could enter the rivers through poor treatment processing (Schmidt 2013). Further, drilling firms were caught illegally dumping fracking water in local streams, ponds and rivers to avoid expensive treatments (Shankman 2010). In these impacted water basins, such water impacts had immediate effects on local ecosystems and populations and eventually could have extensive effects on distant biotic ecosystems and human social systems. As these concerns proliferated in 2009, 2010 and 2011, environmental groups and others concerned with watershed issues geared up to understand the impact of fracking activity, disseminate information on these issues to members and interested citizens and determine potential courses of organizational action. Mobilization on fracking required immediate action that posed challenges to these organizations.

In less than 3 years, between 2008 and 2011, Marcellus activity had gone from a virtually non-existent environmental problem to a widespread, intense and widely impactful environmental burden affecting people and ecosystems across half of Pennsylvania. It was also a problem which cut across many social, economic, political and environmental aspects of community life. Individuals were concerned about well water, local governments about water treatment safety, fishers about recreational impacts, farmers about dangers to livestock and medical personnel about public health effects. The rapidity of the onset of this issue posed real challenges to organizations directly or indirectly concerned with water quality, as well as its impact on the biotic environment and on constituent communities. Most groups were local, concerned with issues in their immediate community, and most were engaged in other projects. Gathering relevant information on the Marcellus water impacts required intense and widespread information seeking activity by organizations with limited resources. Further, public interest in these issues soared as communities and individuals were asked to permit drilling activity. These organizations, often with mandates to act as stewards of watersheds, rivers and natural areas, responded to citizen and constituent concerns by gathering and disseminating information on these issues.

11.3 Rise of Organizational Networks Across Space

With limited resources, a limited time frame and a wide geography of concern, many of these organizations reached out for assistance to any other organization that could assist them in their new efforts. For many, this meant developing new contacts with far-flung organizations and with organizations which they had not previously partnered. This network formation represented both a practical problem for organizations and a theoretical example of the role that networks play in resource mobilization in response to social problems. On a practical level these connections evolved as a strategy for meeting the need for exchange of information and the resources across the Marcellus region of Pennsylvania. Networked groups are able to collaborate by "pooling their knowledge and resources to 'solve a set of problems which neither can solve individually'" (Belaire et al. 2011, p. 472). Once established, these connections coalesced into a more regular network of information exchange, mutual support and for dissemination of information across space and constitute a significant resource for organizations working on fracking issues. Understanding the structure of these networks across space illuminates both a theoretically interesting and practical aspect to the mobilization of the Marcellus Shale environmental advocacy movement. On a practical level, understanding how communities are linked together across space by these organizations helps us understand the strengths and limitations of this network as a resource for collective problem solving. Such networks may be marshaled to enhance purely local institutional resources in resolving community issues (Lyson 2004; Tolbert et al. 1998).

On a theoretical level, the rise of these networks across space adds an important dimension to McAdam and Boudet's approach to "communities at risk for mobilization" (2012, p. 2). Rather than focusing on the social movements themselves, they focus on community characteristics that encourage emergence of collective action. By using communities as the focus for analysis, their approach turns toward the analysis of communities in explaining collective action: "Are there factors—of our communities, their histories, or the specifics of the projects—that help us understand the variation in emergent action that we see in these locales?" (McAdam and Boudet 2012, p. 25). In their analysis of environmental action in 20 communities, they demonstrate the influence of five dimensions of community context: environmental risk (threats to public safety, health, environment and quality of life); demographic composition (income, unemployment, education, home ownership and value); economic base (operationalized as a community's history of similar activities); political engagement (voter participation, in national and local elections); and organizational/civic capacity (the prevalence of non-profit organizations).

Perhaps because these were case studies, the authors did not consider inter-areal influences such as the Marcellus environmental network. Certainly the accumulation of these contacts from dispersed areas is part of the community's organizational and civic capacity. Network structure and the spatial patterns of these networks can add or detract from collective action for a community's constituent population. Indeed, the spatial diversity of contacts increases the resources available to a community that can set up unlikely alliances among diverse groups—an important dimension for the success of a social movement (McAdam and Boudet 2012).

The omission of these spatial networks as an element of civic capacity can be understood given the focus on the internal organization of these movements and generally on micro-level analysis in the social movement literature. Certainly analyzing networks among individuals has a long standing in the collective action literature (Diani 2003b, p. 7). Specifically, these networks have become understood to be especially efficacious for individuals in civic and political action (Baldassarri and Diani 2007; Lake and Huckfeldt 1998); among organizations (Blau and Rabrenovic 1991; Glanville 2004); and encompassing important elements of geography and space.

> We do not want to overstate the degree to which social space is independent of geographic space. We simply want to argue that the key to understanding physical space is to appreciate how it comes to be occupied by complex and dense sets of social spaces. It is much easier to evolve a new social space if one is in direct physical contact with other people who have the knowhow and tools it takes to help found the new social space. The current view of urban agglomeration implies that the creation of new social space is likely to be concentrated where lots of firms, industries, educated people, and government are located (Arthur 1988; Krugman 1991). These actors learn from each other, compete with each other, and are able to produce new, nearby social spaces as they figure out how to take advantage of the opportunities to do so. Indeed, the growth of cities is one of the forces that clearly are involved in the proliferation of strategic action fields. (McAdam and Fligstein 2012, pp. 63–64)

However, treating community position in larger macro-level networks has not been central in this literature. McAdam and Fligstein recognize the importance of these connections among areas as an under-addressed but important dimension. Perpetuating a notion of community that stretches back to the early twentieth century (Galpin 1915; Ravenstein 1889; Zipf 1949) they conceptualize community as comprised of a series institutional action fields, that overlap in space. The authors conclude:

> Probably, the most important issue raised by our perspective concerns the linkages between fields. This is an area that has not been explored empirically very much. The main way in which scholars who study strategic action fields work is to isolate a particular strategic action field, define its relationship to nearby strategic action fields, and then proceed to an

> account of the formation or transformation of a given strategic action field. But the dynamic linkages between strategic action fields have rarely been explored by scholars. But having said that, we need to design studies that look at these relationships over time to see how they produce change and stability in the players in strategic action fields. We know almost nothing about these processes. (McAdam and Fligstein 2012, p. 220)

In their work, these authors have moved the study of social movement into the realm of networks and tied these networks to space, geography and community. Such a focus on social interactions across space has been a lacuna in the collective behavior literature. However, the study of inter-areal interactions has a rich tradition in other sub-fields of sociology as well as demography, human geography, and economics. These approaches have all highlighted the notion that community is comprised of complex fields of interaction and that these interactions are influenced by space (cf. Hawley 1986, p. 50–51; Kaufman 1966; Porter and Howell 2012, Chapter 3; Wilkinson 1970). In these traditions, the exploration of mutual effects between social and physical space is often the raison d'être for research. These spatial sub-disciplines have accumulated a rich set of empirical findings, models and middle range theory about socio-spatial interaction. In this chapter we bring several of the most wide-spread models of socio-spatial interaction to bear on the issues raised by McAdam and others. Specifically, we treat network structure as one aspect of organizational capacity of local communities. This aspect is found in institutional ties to other areas.

As these networks expand in space, the scope of shared resources can also be expanded, thus enlarging civic capacity. This geographic spread can create ties among advocacy organizations that, in turn, help facilitate information diffusion and provide resources, including: administration, staff, member and volunteer time; financial resources; joint participation in specific actions; and shared linkages to third party public and private organizations (Diani 2003b). Where these network connections are thin or non-existent, there are holes in this advocacy network. Our intent is to evaluate the spatial dimension of both the theoretical and practical dimensions of this network response. Our work identifies these network holes and clarifies why they remain and how they might be filled through better networking among environmental advocacy organizations. The role that space plays in this structure is, of course, central to understanding why some communities enhance their civic capacity through these networks and others do not. On a theoretical level, this spatial pattern of contacts can be seen as one dimension of a broader phenomenon of collective action and the development of a nascent social movement. Toward these ends, this study is concerned with evaluating the structure of those contacts across space, understanding where such contacts did and did not occur.

11.4 Scope of Impact: Demographics of Marcellus Watershed Areas

In this section we are interested in the potential impact of this activity on socio-economic population groups. Our interest here is to assess which social and demographic groups are most likely to bear the brunt of such impacts. To do this we need to establish the geographic/environmental spatial scope of impact. As discussed above, environmental and health burdens associated with fracking impact first and foremost through the network of streams and rivers. Land (and people) may be grouped into cohesive geographic units (watersheds), where all of the water that is under it or drains off of it goes into the same place (US Environmental Protection Agency [EPA], 2013). Because the flow of rivers, streams and aquifers link populations without regard for political and administrative boundaries, water related impacts on populations are best analyzed in reference to their common watershed. Settlement patterns and ongoing community boundaries are shaped by these watersheds (EPA 2013). In general, water tables (aquifers) conform to surface watersheds and represent common water for human consumption. The quality of that water directly affects the health of the constituent population.

For these reasons, watersheds are often explicitly or implicitly the geographic focus for government and political discourse. Management of watersheds for health, recreation, industry and the public good links otherwise geographically separate municipalities, states and countries. This is the case in Pennsylvania. The State Water Plan utilizes watersheds as units for many of its water and environmental impact assessments (Environmental Resources Research Institute 1998). The state is divided into three major river systems, then 20 smaller tributary drainages (sub-basins) and finally by minor tributaries of which there are 104 (watersheds) that vary from 100 to 1000 mile^2 (Pennsylvania Spatial Data Access 2013).[1] Any potential Marcellus impacts can readily spread within these watersheds, but are relatively contained between these watersheds, making these geographies ideal for describing populations at risk. Watersheds comprise an important geographic boundary for populations. They link apparently spatially dispersed populations by common environmental factors and, in this case, by common environmental impacts.

However, to assess the possible impacts on populations, it is necessary to map the naturally occurring watersheds to the administrative boundaries used by the US Census. That is, we wanted to match the 104 watersheds with constituent census information using census reporting units. Most units (cities, counties, places and even census tracts) simply cross the boundaries of these watersheds. This mismatch makes these units problematic for assessing spatial impacts and for identifying affected populations (Irwin 2007). Block groups provide the tightest census geography with detailed population characteristics that could potentially be used to describe watershed demographics (US Census Bureau 2007a).

Using ArcGIS, 80 % of the block groups were mapped directly to watersheds. The remaining block groups, virtually all at the edges of each watershed, were assigned to the 104 watersheds using the smaller census block geographies. Block groups were assigned to the watershed where the greatest percent of the population resided. If this assignment was still ambiguous, the block group was assigned based

[1] Pennsylvania's 104 State Water Plan watersheds subdivide the 58 watersheds identified by the EPA in the State along hydrological, social and political divisions.

upon its downstream location relative to the watersheds in question since it was the upstream watershed directly affecting the constituent population. The result was to create census unit-based watersheds that oriented the constituent population to the natural watershed. The final categorization was evaluated by comparing categorization of census blocks under the assigned watershed, compared to the categorization of the blocks using direct centroid matches to the watershed (US Census Bureau 2007b). Differences may be regarded as the degree of error in assignment. We examined both in terms of population and in terms of area under the two watershed categorizations: natural and census watersheds. There was a 3.5 % absolute difference in area under the two schemes and a 2.6 % absolute difference in population under the two schemes. For our purposes, evaluating population characteristics of the watersheds, this would not constitute a major source of bias.

Having established the spatial relationships between the natural areas and demographic areas, we can proceed to assess similarities and differences in Marcellus and non-Marcellus population characteristics. Figure 11.1 shows the Marcellus region buildup from census watersheds. Every census watershed with at least one Marcellus permit is included. The resultant geography shows a region stretching in a band from southwest through northeast Pennsylvania that encompasses the prime Marcellus Shale geology. This region contains the population in census year 2000 grouped by watersheds that would be exposed to Marcellus drilling. What were the demographic differences between the Marcellus and non-Marcellus regions when drilling began? Table 11.1 uses Census 2000 demographic characteristics to answer this question.

The Marcellus shale region population tends to be less minority, more rural and in lower density area, with fewer young children, lower household income and fewer white collar workers. Household income is more concentrated under $200 k in the Marcellus region. Having established a few basic demographic divides between the affected populations in the Marcellus regions and the rest of the state, the next question is whether there are significant differences in populations in areas of high Marcellus activity versus low impact activity, within the Marcellus region. Table 11.2 shows that there are important differences.

Within the Marcellus region, watersheds with higher proportions of rural populations, farmers, and young children tend to be the areas with more drilling. Areas with more drilling also tend to be watersheds with more households concentrated at lower incomes.

At the beginning of the Marcellus build up, those areas that were most directly involved in individual and community decision making tended to be white, rural poor farmers in areas without a wealthier income strata present. Those areas with the highest levels of permit activity (and drilling) tended to be more rural, more likely to be farmers, even poorer than the low impact Marcellus watersheds, with few wealthier people and with more young children (a group especially vulnerable to health risks associated with environmental drift of chemicals). Thus the concentration of Marcellus drilling activity would be in communities with the greatest economic needs. As leasing and drilling boomed, these populations could potentially look at tremendous personal or community windfalls. On the other hand, these

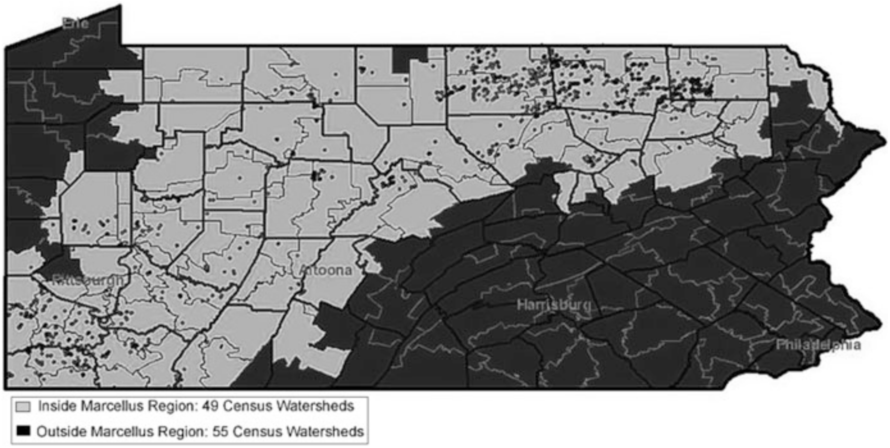

Fig. 11.1 104 census based watersheds & Marcellus well permit activity through 2009

Table 11.1 Demographic differences between Marcellus and non-Marcellus watersheds, from census 2000

Comparisons inside (N = 49) and outside (N = 55) Marcellus region for 104 census watersheds		
	Marcellus region	Mean
% Nonwhite (only)	Outside Marcellus Region	8.24
	Inside Marcellus Region	3.38**
% Hispanic population	Outside Marcellus Region	2.79
	Inside Marcellus Region	.64**
% Rural	Outside Marcellus Region	44.06
	Inside Marcellus Region	58.02*
% Rural pop that are farmers	Outside Marcellus Region	3.07
	Inside Marcellus Region	2.32
Pop per square mile	Outside Marcellus Region	595.27
	Inside Marcellus Region	215.23*
Median household income	Outside Marcellus Region	$42,188
	Inside Marcellus Region	$34,507**
Ratio household income <$200 K to >$200 K	Outside Marcellus Region	11.28
	Inside Marcellus Region	14.40**
% Population 10 years & under	Outside Marcellus Region	63.53
	Inside Marcellus Region	62.24**
% Workforce in manufacturing	Outside Marcellus Region	11.36
	Inside Marcellus Region	10.78
% Workforce in white collar occupations	Outside Marcellus Region	41.39
	Inside Marcellus Region	36.62**

*Mean difference significant at the 0.05 level
**Mean difference significant at the 0.01 level

were populations making a living from the land along with cultural activities focused on hunting and fishing in local streams and forests. These were groups who could find their ways of life threatened by Marcellus impacts.

Table 11.2 Demographic differences within Marcellus drilling region, from census 2000

Correlations between drilling activity & selected variables in 49 census watersheds	
	Marcellus permits per 1000 persons
% Nonwhite	−.168
% Rural	.492**
% Rural pop that are farmers	.384**
Pop per square mile	−.212
% Hispanic	−.079
% Children 10 & under	.333*
Median household income	.185
Household income <$200 k to >$200 k	.315**
% Manufacturing	.252
%White collar	−.205

*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level

As these communities sought clarifying information, one might expect them to turn to the local organizations and institutions that had always been most concerned with watershed issues—sportsmen's groups, environmental organizations, county conservation districts, and local state parks—for further information. These groups might welcome overtures from more distant organizations as a way to enhance access to information and to meet local interest. Such networks connections could enhance the organizational capacity in these communities. On the other hand, the demographic characteristics of these areas were much like the community factors identified by McAdam and Boudet as resistant to mobilization (2012, p. 182). They were poor, in areas with few civic organizations. These were low density areas with lower property values and a conservative rural culture. Most of these communities had long experience with primary resource extraction and had intergenerational history with mining, gas extraction, and lumbering. Unemployment was high and economic need great. As Marcellus activity geared up, one might expect these factors to moderate against network mobilization.

11.5 Environmental Networks

Data on the structure and character of the environmental organizational network was gathered by Pischke (2013) in the summer and fall of 2012. Survey information includes data on the spatial dimensions of organizations, the number and character of network linkages among environmental advocacy organizations concerned with Marcellus hydro-fracturing, and additional data on frequency and character of exchanges among network partners. The population of interest was all self-identified environmental organizations that have non-profit status and stated missions to protect or conserve the natural environment.

Counties were used to identify the location for organization connections and activity, which in pretesting and in practice, were the most salient geographic identifiers for organizational representatives.[2] In this case, organizing spatial information by county made the most sense. The activities associated with these organizations are often bounded by county laws, mandates, etc. Samples were drawn from two regions in Pennsylvania—a group of twelve counties in the northeast and another group of eleven counties in the southwest. These regions were defined as planning regions by Citizens for Pennsylvania's Future (PennFuture, a statewide environmental advocacy organization). We included the following counties in northeastern Pennsylvania: Bradford, Carbon, Lackawanna, Luzerne, Monroe, Pike, Potter, Sullivan, Susquehanna, Tioga, Wayne and Wyoming. Southwestern Pennsylvania included Allegheny, Beaver, Bedford, Blair, Cambria, Fayette, Fulton, Greene, Somerset, Washington, and Westmoreland Counties.

These two regions reflect "diverse arrays of demography: everything from extremely sparsely populated forests to major metropolitan cities" (Jacquet 2009, pp. 51–52). They also represent distinct histories, industrial bases, settlement patterns, and political orientations—elements identified by McAdam and Boudet as foundational differences in collective action (2012, p. 40). Further, they neatly break up interactions into short and long distance. Bridging beyond the locality is vital, and we also include organizations that are part of the wider state and national environmental movements (Bryan 2004, p. 882).

The loose ties that link these distinct theoretical categories into larger networks are critical dimensions for resource mobilization (Carmin 1999, pp. 101–103). They provide coordination, effective outreach, education, communication and design of legislative/legal remedies (Bullard and Johnson 2000, p. 561; Snow and Soule 2010, p. 88). Overall, using two geographically distinct study areas provides two categories of social and spatial heterogeneity that might be expected to influence network formation among environmental organizations and are meaningful in presenting results.

The contact information for public representatives was obtained using Guidestar, an online service with information on non-profits' mission, finances, staff, and board (www.guidestar.org). Only those organizations with experts willing to report on their activities were included in the survey. Organizations were excluded if such experts were not available. After initial telephone contact with the spokespeople, they were asked directly for an email address where they could receive the questionnaire or, when necessary, they were asked for the phone number and email address of someone better able to complete the questionnaire. In some cases, the representative contacted completed the questionnaire and also provided another person in their organization who could also do so.

[2] Pischke (2013, p. 98) also attempted to have organizational activity identified by watershed area. Although many organizations were familiar with their own watersheds, they were seldom familiar with specific watersheds for partner organizations.

Each organization's representative was emailed a brief explanation of the study, instructions for completing the questionnaire, a link to the online questionnaire and a date by which they should complete it. These public spokespeople had the knowledge and capacity required to answer questions about their organization's Marcellus Shale activities (such as the president, manager, public relations employee, outreach coordinator). Representatives from 123 organizations in southwestern and northeastern Pennsylvania were asked to participate in the questionnaire. Fifty-eight of these organizations completed the questionnaire, for a response rate of 47 %.

The network generator questions in this survey use the approach developed by Belaire et al. (2011), where each respondents chooses from a list of environmental organizations operating in each county. The respondent is then asked to give information on the types of collaboration they perform with specific groups (Belaire et al. 2011, p. 466). When these network partners were not one of the core 58 surveyed groups supplemental information was added from the organization's webpage from respondents, interviews with key informants and through Internet searches, per Belaire et al. (2011, p. 468). Survey responses from the 58 core organizations identified an additional 138 organizations that partnered with the core group on Marcellus issues, bringing the number of organizations in this network analysis to 196. These additional organizations covered most of the originally targeted organizations that did not return surveys.[3] Many of these partners were readily identified as falling into the three original categories (sportsmen's groups, county conservation districts and local conservation chapters). However, many of these environmental groups regularly consulted with representatives from the state parks in Pennsylvania on Marcellus issues as well. We have included state parks as a fourth category for network exchanges among these environmental groups.

Figure 11.2 shows the geographic locations of these organizations within county, as well as the two study regions. Fifty-nine percent of our identified organizations were in the southwest region—most concentrated in Allegheny County (where Pittsburgh is located). Forty-one percent of our study organizations were in the northeast and were relatively evenly distributed across the region. Organizations with environmental protection as their primary mission comprise 44 % of the study organizations, sportsmen's organizations were 20 % and county conservation districts another 14 %. State parks—widely identified as important in the networks of these three types of organizations—make up the remaining 22 % of organizations in the study.

From this network information, we used network analysis and GIS techniques to map the spatial coverage of environmental organizations to determine the degree

[3] Since our informant organizations were able to identify groups they worked with throughout the state, the network of target or destinations for these groups are the more complete network (196 potential organizations as opposed to 58). In subsequent network analyses we will use destinations to calculate the measures of network structure.

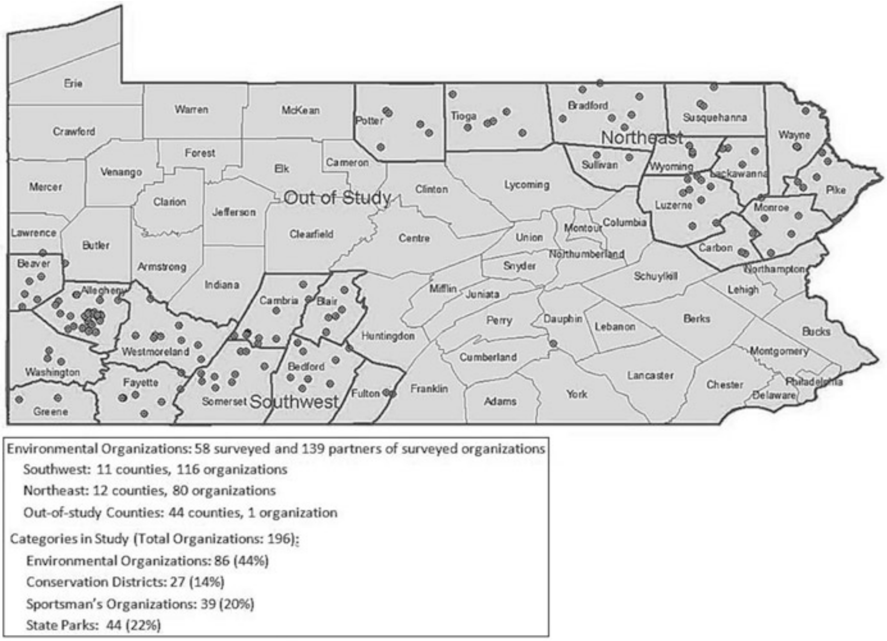

Fig. 11.2 Geographic distribution of environmental organizations & study areas

and extent of environmental protection afforded by these organizations. Figure 11.3 provides counts of these organizations by county and organizational ties among counties. Each line represents an aggregate number of ties between each pair of counties. Thus, this map represents the spatial flow of organizational contacts from county to county.[4] Below we discuss the organization character and content of these flows.

11.6 Organizational Resources & Activities: The Content of Network Exchange

The missions of these organizations represent diverse environmental concerns. Few of these organizations had broader missions that prioritized larger territories of activity. Twenty-two percent of organizations focused on environmental justice, 18 % conduct research, and 14 % lobby politicians. Only a handful of those respondents' organizations' missions focused on public health, or general

[4] Additional information on geophysical conditions, hydro-fracturing well location and geographic distribution of well permits were derived from Pennsylvania Department of Environmental Protection reports.

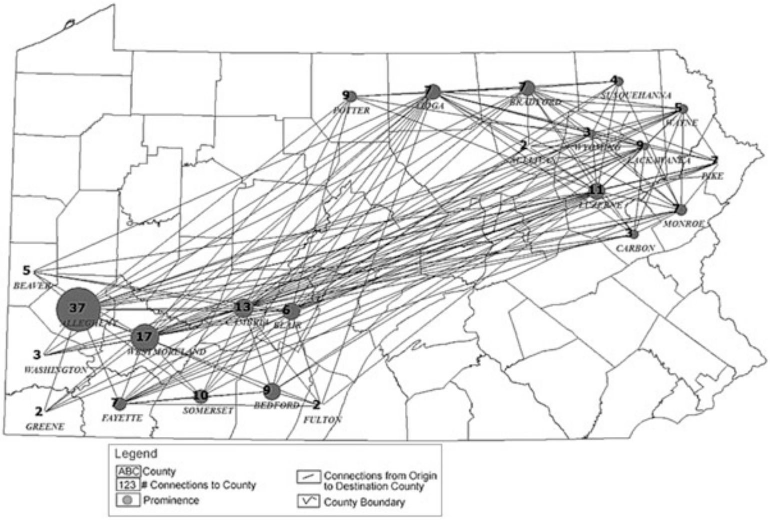

Fig. 11.3 Geographic distribution of 185 organizations with at least one network tie

conservation and preservation of natural resources. Most groups were formed for direct action on specific lands or in defined territories. Seventy-two percent of these groups had a primary focus on restoration habitats and 68 % had missions dedicated to work on land conservation. For these groups, concern was vested directly with natural resource conservation and only secondarily with social and economic dimensions of the environment. Most of these organizations (76 %) operationalized their mission through public outreach and educational activity. That is, most saw the gathering and dissemination of information as central to the performance of their missions, although for most, this was a local audience. In many ways, the rise of Marcellus activity changed this local orientation.

The majority of the organizations began working on Marcellus Shale issues, at least internally, in 2008 when drilling began in Pennsylvania. None of the representatives interviewed indicated that their organizations have been working on the issue before 2006. Two organizations only began to focus on gas drilling as a secondary issue in 2011. Notably, only one did so because they were approached by other organizations to represent them. Whether through influence of increased interaction among environmental organizations or through direct threats to their stewardships of land, concern with Marcellus activity became prominent for most of these organizations.

In this survey of 58 environmental groups, every organizational representative indicated some level of concern with Marcellus Shale issues. Thirty-four questionnaire respondents, or 69 %, indicated that their organizations are "very concerned" with Marcellus Shale drilling activities. Virtually all of the groups with environmental stewardship as part of their mission saw the Marcellus impacts as of utmost concern to their mission. The organizations that expressed lower levels of concern include county conservation districts, which liaise with gas companies to grant permits or monitor drilling, and sportsmen's groups that have negotiated or received royalties for gas extracted from their properties or game lands. Several organizations indicated that they do not focus on Marcellus Shale issues or have remained neutral on the subject because they do not have wells in their county or are not otherwise directly affected by this industry. Even for these organizations, Marcellus issues clearly emerged as a group concern of sufficient import to have developed a position.

In less than 5 years (2006–2011) Marcellus drilling issues engaged all these organizations and most saw this new activity as a direct threat to their missions. Further, this threat cut across the diversity of concerns among these organizations. One reaction to this threat to mission was to shift organizational activity toward Marcellus issues. It is not surprising that the approaches used by these groups built on past expertise gathering and disseminating information. When organizations work on Marcellus Shale drilling issues, the top priority for 78.3 % of groups is cultivation of public awareness through environmental education, educational seminars or public webinar, followed by 45.7 % of organizations working on monitoring existing legislation or policy implementation. Direct action on Marcellus is far less common. Between 19.6 and 28.3 % of organizations perform direct actions, serve on an advisory committee, formulate new policies or regulations, and conduct research, whereas 8.7–17.4 % of organizations develop legal strategies, lobby congress, state legislatures, county boards of supervisors or municipal councils as well as international, federal, state, or local agencies.

Much of this activity required shifting the time, effort and resources within existing organizations—utilization of resources that clearly taxed the resources of these small, often volunteer, groups. Additionally, and perhaps because of this, these groups became much more active in seeking out partners for sharing these burdens. When asked about the nature of collaboration with other environmental organizations, many group representatives stated that they partner with others for direct action. Fifty-eight percent of organizations report collaborating with other groups on projects related to Marcellus Shale activities. A little less than half of the organizations, or 44 %, monitor existing legislative or policy implementation, 37 % serve on advisory committees, 27 % formulate new policies or regulations, educational materials or research, 13 % lobby congress, state legislatures, county boards of supervisors or municipal councils, and 3 % lobby international, federal, state or local agencies.

As organizations expanded their activities on Marcellus drilling issues beyond their previous scope of work, many of these groups reached out to others to share the increased organizational burden of direct action. Sharing resources for direct action is clearly advantageous and an impetus for developing network partnerships. But it is information sharing across organizations that became one of the driving forces bringing these organizations together. When asked about collaboration with other organizations on Marcellus drilling activities, information sharing was nearly universal in the network. The most common interaction among all organizations, or 93 % of them, involves the exchange of ideas. Dissemination of information to the public follows closely behind. Sixty-nine percent of organizations collaborate on cultivation of public awareness.

The character of exchange is different for information flow than for material exchanges of resources. The latter (resources) are limited forms of exchange typically directly shared only by two organizations in the network. These dyadic resource exchanges can include monetary expenditures on direct action, sharing personnel, mobilizing membership time for mailings, and similar resources that are no longer available to others organizations once committed to a partner organization. This zero-sum character means that network flows of resources tend to be dyadic—benefiting direct exchanges among partners. More partners mean more potential dyadic exchanges. Of course, information flows also occur in these dyadic exchanges; however, expressing access to information only in dyadic terms understates an organization's access. Indirect flows of information are also important.

Unlike material resource exchange, information exchange is not intrinsically a zero-sum transaction. Information shared with a partner is not lost, and pooling information and ideas does not deplete a stock of ideas. This means that exchanges of information are readily transmitted across the network both directly to partners and indirectly to the organizations that those partner's contact. Thus access to information is determined directly (number of contacts with other organizations) and also indirectly—the number of partners those organizations contact (second tier) and so on (third tier, fourth tier, etc.). This gives us two conceptions of network advantage for organizations and places—one based on direct ties and most relevant to resources exchanges, and the other encompassing all ties in a network and most relevant for information dissemination and exchange. Measurements of these types of network position are termed centrality. Below we calculate two measures of centrality developed by Bonicich (1987) for measuring flows across networks and applied to networks among places (Irwin and Hughes 1992; Hughes 1993).

Table 11.3 provides the centrality score of each county in this network based on direct exchanges among these environmental organizations. Values above one reflect greater than average centrality (or number of partners) while those with fewer than average direct ties are less than one. For each county, scores are given for both the observed network and the hypothetical network where all organizations in a county are connected. This later measure represents centrality in a hypothetical network where all environmental organizations working in a county are exchanging information and sharing resources. Differences between scores signify departure from a centrality based on a geographic (county) maximum.

The observed network in the northeast reflects a relatively even distribution of organizational ties among these counties. No one county dominates the network in this region. Luzerne County is the most prominent area in the northeast (with a centrality score of 1.8) and is more central than one might expect given the number of organizations working in that county (centrality of 0.95). That is, Luzerne has more direct connections with organizations within and across these regions than one would expect from the number of organizations in that county. This concentration of network connections provides a structural advantage for this area as an accumulation point for resource and information exchanges. It also reflects the influence of this county on other Pennsylvania counties in this study. In the northeast, information and resources flow from Luzerne. To lesser degrees this pattern of influence and concentration of network ties is found for Tioga, Bradford and Lackawanna Counties. Most of the counties with less than average observed

Table 11.3 Observed and maximum connections, centrality based on direct ties

	Observed connections, network centrality	Maximum possible connections, network centrality	Percent deviation from maximum
Northeast counties			
Bradford	1.23	1.07	17 %
Carbon	0.56	0.59	−3 %
Lackawanna	1.14	0.71	43 %
Luzerne	1.80	0.95	85 %
Monroe	1.09	0.83	26 %
Pike	0.42	0.59	−17 %
Potter	0.61	1.19	−58 %
Sullivan	0.13	0.47	−34 %
Susquehanna	0.59	0.59	−1 %
Tioga	1.48	0.83	65 %
Wayne	0.77	0.95	−18 %
Wyoming	0.87	0.59	28 %
Southwest counties			
Allegheny	2.96	3.81	−85 %
Beaver	0.10	0.71	−61 %
Bedford	0.88	1.19	−31 %
Blair	1.47	0.83	65 %
Cambria	1.75	1.54	20 %
Fayette	0.58	0.71	−13 %
Fulton	0.39	0.23	16 %
Greene	0.07	0.23	−16 %
Somerset	0.71	1.31	−60 %
Washington	0.12	0.71	−59 %
Westmoreland	3.28	2.38	90 %
Network average:	1.00	1.00	

access to the network are actually near their maximum centrality scores—that is they are about as connected as could be expected given the limited number of environmental organizations working within their boundaries. Along with Carbon County, Potter and Sullivan Counties have the lowest centrality scores. These latter two also have considerably fewer network connections than might be expected given the number organizations in these counties.

All three counties are also relatively isolated in this network, as shown in Fig. 11.4. This figure is a sociomap that clusters these units based on closeness or network distance (de Nooy et al. 2011, p. 146). The core counties (Luzerne, Tioga and Lackawanna) occupy the central social space in this network. Carbon County is

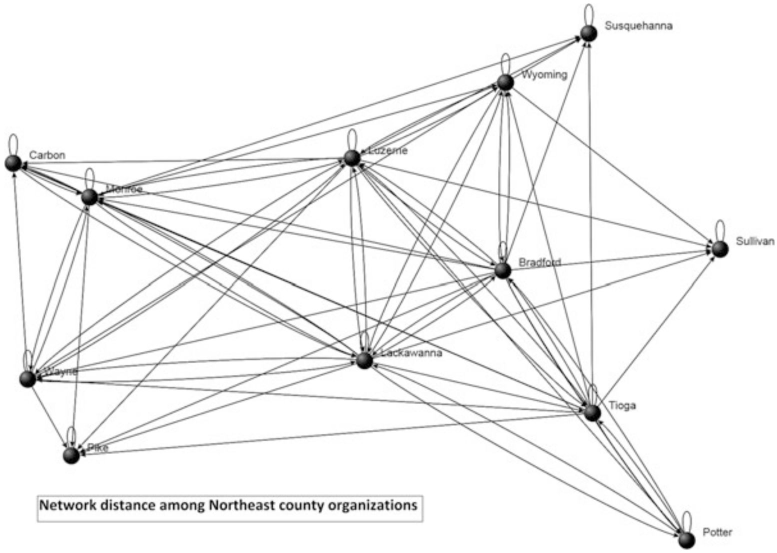

Fig. 11.4 Sociomap of the Northeast network

most distant (in network connections) from these core areas, although this area is integrated into a regional network with Monroe, Wayne and Pike Counties. Sullivan and Potter Counties are, however, relatively distant in network space from the other areas, indicating a peripheral position in the structure of organizational contacts. These counties are more socially isolated in that they are less connected in the environmental exchange network. As noted earlier, both areas also are far below their potential for integration into this network. Simply, linkages from the constituent populations of Potter and Sullivan Counties to the broader advocacy network are weak.

As seen in Table 11.3, the southwest region exhibits a much more hierarchical pattern than the northeast region. The southwest region has two extensively dominant centers (Allegheny and Westmoreland Counties), two moderately dominant counties (Blair and Cambria) with the remaining seven counties clearly subdominant. The two dominant counties did have the largest number of organizations of any places in this study. Allegheny had 32 and Westmoreland 20 environmental organizations; however, the Westmoreland organizations were more extensively connected, leading to a higher centrality score than Allegheny and a centrality score 90 % higher than expected. Similarly, Blair and Cambria Counties were more central in this network than might be expected. Three counties had very low centrality levels (Beaver, Greene and Washington). While Greene might be expected to be marginal in these networks given the limited numbers of organizations working in this county, Beaver and Washington both have far fewer connections than might be expected given the environmental organizations in those areas. Again, a social map of these connections, Fig. 11.5, depicts all three areas as

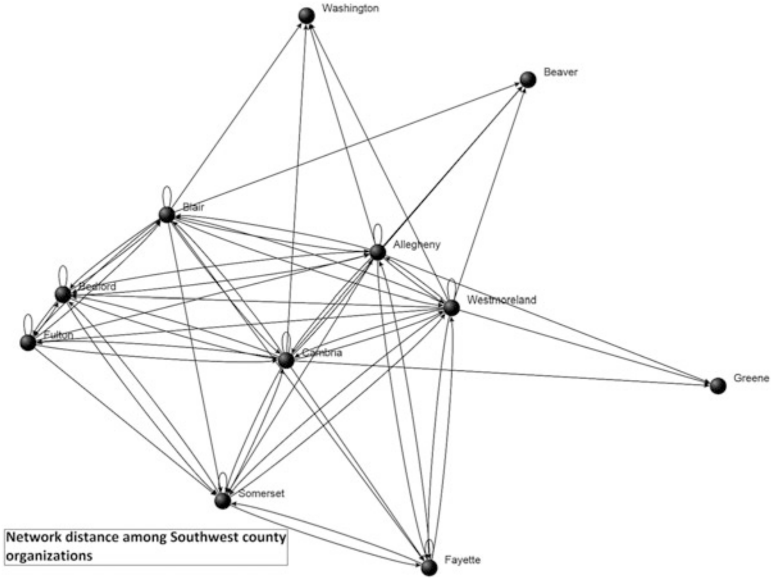

Fig. 11.5 Sociomap of the Southwest network

socially distant from the rest of the regional network. In all three cases, as seen in the centrality scores, these low scores represent deficits of organizational ties that could well hamper potential access to resources and information—especially the former.

Information access, as discussed previously, requires a more complex formulation of access and centrality. Table 11.4 does this. It provides centrality scores based on the direct and all indirect organizational connections among counties. In this calculation of centrality, the number of direct connections increases the centrality score (as in the last table) but being connected to counties that are also central in the network increases the score (unlike the last table). Thus, an organization or place has better access to information when it has more connections, but that information access is enhanced when the partner is also widely connected to other organizations (Irwin and Hughes 1992). Centrality here also may be thought of as reflecting the relative network "closeness" of any place to all others (Bonacich 1987). Counties with the highest centrality score have both the widest organizational reach and can connect with all other areas through the fewest number of connections.

In general, the pattern of centrality by county is similar to that for direct ties. Centrality for most counties is what one might expect from the number of organizations working in the county. The most central places are not strongly dominant. In terms direct and indirect access to information flow, only Luzerne stands out. Simply, most of the organizational connections in these areas are not to prominent

Table 11.4 Observed & maximum connections, centrality based on direct and indirect ties

	Observed connections, network centrality	Maximum possible connections, network centrality[a]	Percent deviation from maximum
Northeast counties			
Bradford	0.94	1.07	−12 %
Carbon	0.47	0.59	−12 %
Lackawanna	0.95	0.71	24 %
Luzerne	1.41	0.95	46 %
Monroe	0.82	0.83	−1 %
Pike	0.37	0.59	−23 %
Potter	0.60	1.19	−59 %
Sullivan	0.14	0.47	−34 %
Susquehanna	0.35	0.59	−25 %
Tioga	1.09	0.83	26 %
Wayne	0.58	0.95	−37 %
Wyoming	0.72	0.59	13 %
Southwest counties			
Allegheny	4.40	3.80	60 %
Beaver	0.21	0.71	−50 %
Bedford	1.01	1.19	−18 %
Blair	1.34	0.83	51 %
Cambria	1.72	1.54	18 %
Fayette	0.76	0.71	5 %
Fulton	0.38	0.24	14 %
Greene	0.14	0.24	−10 %
Somerset	0.93	1.30	−38 %
Washington	0.26	0.71	−45 %
Westmoreland	3.43	2.37	105 %
Network average:	1.00	1.00	

[a]In a maximally connected network the relative centrality will be the same as in the direct ties network

places. Potter and Sullivan, the two counties that had the greatest departure from maximum in terms of direct ties (significant for resources), are also underperforming in terms of indirect ties (information). In the southwest, Allegheny and Westmoreland Counties remain clearly dominant areas, maximizing access to both information and resource flows throughout the regions of study. Notably, the rank of these two counties are reversed with Allegheny clearly exhibiting both the widest spread of network connection and “closest” to all other areas through these organizational connections. This is, of course, a significant advantage as an accumulation point for the spread of information.

The structure of network connections seems largely due to the pre-existing numbers of organizations found in these counties. However, there are exceptions. As before Beaver, Greene and Washington Counties have the lowest scores. In this sense these three subdominant areas are the most ‘distant’ counties in the network, despite geographic proximity to the centers of these networks—Allegheny and Westmoreland Counties. Why might this be? To understand the role of geographic proximity and why network connections may (or may not) follow a distinct spatial pattern, we turn to explicitly spatial models of the connections among organizations across counties.

11.7 Four Models of Spatial Interaction

Hawley observed that because humans are necessarily limited in time we are also limited in space (Hawley 1986, p. 5). For this reason, the friction of travel imposes a time constraint across space that increases the burdens of interaction as distance increases. As Tobler famously stated, “Everything is related to everything else, but near things are more related than distant things” (1970, p. 236). Of course, modern transportation has moderated this friction of distance and communications have superseded time barriers to information exchange. Still, distance manifests costs to travel, limits face-to-face contact, and provides challenges to coordination among organizations. Social divisions and cultural differences can also increase with distance so that even where physical barriers to interaction are minimal there are distances beyond which regular daily social interaction are limited (Irwin 2007). Some of the earliest formulations in social science began with analysis of the potential draw among places and the role that constraints of distance played (Ravenstein 1889). While this spatial resistance was most obvious in the movement of material objects (people, products, etc.), social science has long recognized that distance also limits intangibles such as the flow of information. Zipf (1949) argued that information seeking will tend to follow the path of least resistance and cease at a point where acceptable results are found. Organizational connections will overcome the frictional constraints of distance when the information (or goods, or jobs, etc.) cannot be met locally. Once established, these information and resource exchanges tend to become routinized. Organizational inertia helps distant connections persevere (Zipf 1949).

These formulations of place, space, and interaction fit the development of connections among environmental organizations. In an environmental crisis, limited resources and information move a county’s organizations to seek contacts with organizations in distant areas. When resources (e.g., numbers of environmental organizations) were not sufficient in an area, these organizations extended their networks beyond county boundaries. These three factors: distance, origin conditions, and destination resources, have been operationalized in social sciences, geography and economics through the use of gravity models. These models have proven to be empirically robust at the aggregate level, if somewhat oversimplifying a complex decision process at the individual or organizational level (Anderson 2011). Nevertheless, if the geographic interdependence of these organizational connections is predictable from the character of places and the friction of distance, then space is clearly playing a role in shaping these networks.

To what extent does this network follow a predictable spatial pattern based on origin and destination factors? To answer this we will first calculate the actual connections across space (Model 1) to estimate typical distance between connected organizations. Simply, this estimates the frictional constraints of distance. We then use this information for three additional hypothetical models of interaction: origin constrained gravity models (de Jong and Van der Vaart 2010 pp. 129–130).[5] In these models the number of connections in the originating county is limited to the number observed in the data. This is the origin constraint. For example, the 32 organizations in Allegheny County generated 125 connections to other organizations. We are assuming that this number of connections approximates the amount of resources or external activity that these organizations could or would devote to partnerships and connections with other organizations.

Figure 11.6 then maps out the geographic distribution of this resource, as allocated by these organizations. What drives this spatial allocation of connections? These models predict that if distance plays a role, then there should be more connections closer than farther. They also predict that some condition or attraction

5

$$\mathrm{Ti\,j} = \mathrm{B\,j}{*}\mathrm{Wi}{*}\mathrm{D\,j}{*}f(\mathrm{Ci\,j}, \beta) \quad (1)$$

$$Bj = 1/(\Sigma i Wi{*}f\backslash(Cij, \beta)) \quad (2)$$

$$f(Cij, \beta) = exp(-\beta{*}Cij) \quad (3)$$

where:

Tij = *the estimated number of organizational connections between origin county i and destination county j.*

i = *the attraction value for destination county j.*

β = *the distance decay parameter*

Cij = *the distance between origin i and destination j*

Formula 1 calculates the actual number of trips in the matrix of origins and destinations. Formula 2 makes sure the total number of trips to the destinations in the matrix is equal to the set number (the origin constraint).

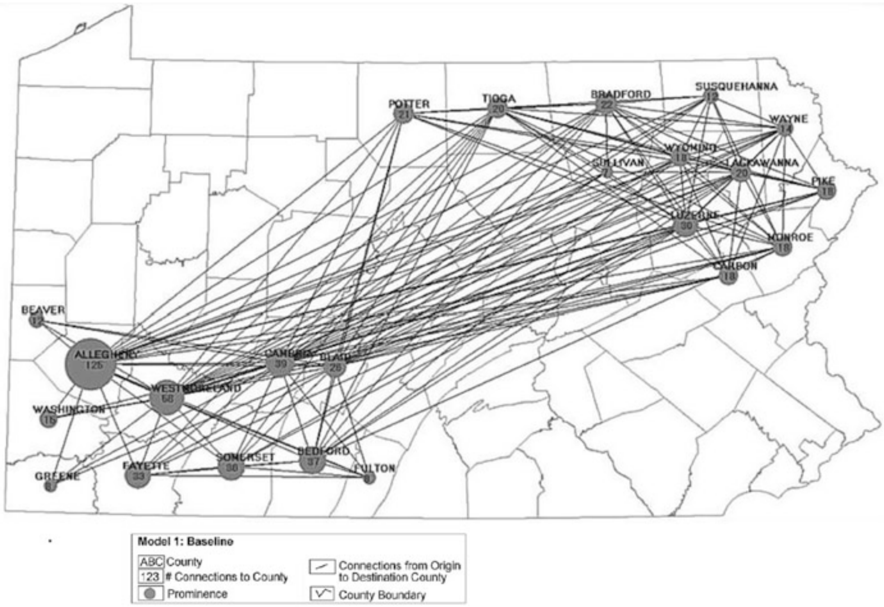

Fig. 11.6 Model 1: Observed connections in destination county

in the destination area can overcome some of the frictional constraints of distance. In Models 2 through 4 we consider three variables that might provide such an "attraction" the number of environmental organizations in the destination county, the amount of activity (as measured by number of external connections) in the destination county, and finally, the number of permits (as of July 2012) for Marcellus drilling in the county. These models all estimate the number of contacts that could be expected in the destination counties (and where those contacts originate) based on destination attraction or draw. In the first two models we use two slightly different measures of organizational capacity as the element drawing contacts from distant areas. The last model predicts the spatial and network distribution of organizational partnerships based on the degree of environmental impact in the destination area, again moderated by distance.

Figure 11.6 shows the number of connections to each county, and specifically, the number of organizations partnering with groups in each county. These destination counts provide a baseline to compare subsequent models. Table 11.5 provides numeric comparisons, while Figs. 11.7, 11.8, and 11.9 provide visual comparison of these models.

Figure 11.7 shows the expected organizational activity destined for each county. Figures 11.6 and 11.7 produce similar networks and similar amounts of activity in each county. As seen in Table 11.5, counties in the southwest region have slightly fewer connections remaining in their region than would be expected in Model 2. Of course the converse is true for the northeast. Overall, organizational interconnections across counties do seem to be driven by organizational activity in those counties, although there is more long distance partnering than one might expect. This effect is relatively modest.

Our alternative model of organizational capacity uses the number of organizations in the destination county to drive connections. This model produces similar results to Model 2. It is a structure closer to the observed patterns (Fig. 11.6). In other words, the pattern of Marcellus environmental connections are most consistent with a gravity model based simply on distance and the number of potential partners in the destination county.

Does the geographic distribution of Marcellus activity shape the spatial patterns in this network? Figure 11.9 maps how this network would look were organizations to concentrate their inter-organizational partnerships in areas where the most Marcellus activity exists. The resultant pattern would reorient the activity to counties now in the periphery of each region—Washington and Greene in the southwest, Bradford, Tioga and Susquehanna in the northeast. Further, as shown in Table 11.5, were fracking activity driving organizational connections, we would expect a significant shift in organizational effort from the southwest to the northeast. Clearly, an environmental risk model of network structure is not consistent with the observed network structure.

11.8 Conclusions

Overall, the spatial pattern of network connections closely follow a spatial model where organizational connections are made on the basis of resources in origin counties (which represents the amount of effort that can be devoted to external contacts) but is limited by distance and the number of organizations in the destination county. The fewer organizations there are in the destination county, the fewer connections will be made. When two counties are of greater distance, the number of connections correspondingly decreases. In one sense, this describes a pattern where the pre-existing spatial distribution of these organizations is the driving force creating the spatial network. As environmental advocacy organizations increased inter-organizational contacts for Marcellus activity, those contacts were made where other environmental organizations were found, when those organizations were not too distant. This pattern of contact does not seem to be affected by the spatial distribution of Marcellus impacts or by the amount of activity these organizations were devoting to an area. In part, this seems to contradict McAdam and Fligstein's (2012) suspicion that social factors might well have overcome the material limitations of geography and that the organizing force in spatial interactions might be driven by pre-existing social elements:

> Our modern conceptions of time and space have been greatly altered by improvements in technology, communications, and transportation that have increased our ability to be aware

Table 11.5 Four models of spatial interaction

Observed and maximum connections, centrality based on direct ties only			Model 1: Baseline	Model 2: Connections driven by organizational activity	Model 3: Connections driven by number of organizations	Model 4: Connections driven by fracking activity
	Orgs. (with connections)	Permits (Cumulative to July 2012)	Observed connections (in destination county)	Origin constrained connections (in destination county)	Organization driven origin constrained connections (in destination county)	Permit driven origin constrained connections (in destination county)
Northeast counties						
Bradford	7	2,327	*22*	19	20	124
Carbon	3	0	*18*	15	8	0
Lackawanna	9	29	*20*	17	25	2
Luzerne	11	16	*30*	27	33	0
Monroe	7	0	*18*	14	18	0
Pike	7	0	*18*	13	16	0
Potter	9	230	*21*	19	27	14
Sullivan	2	210	*7*	6	6	12
Susquehanna	4	1,027	*12*	10	11	47
Tioga	7	1,727	*20*	18	21	96
Wayne	5	20	*14*	18	18	1
Wyoming	3	245	*18*	11	9	14
NE Region total	*74*	*5,831*	*218*	*187*	*212*	310
Southwest counties						
Allegheny	37	51	*125*	137	138	5
Beaver	5	39	*12*	11	16	3
Bedford	9	2	*37*	39	32	0
Blair	6	10	*26*	29	22	1

Cambria	13	26	*39*	45	50	3
Fayette	7	360	*33*	34	24	34
Fulton	2	0	*8*	8	7	0
Greene	2	1,082	*8*	7	6	86
Somerset	10	52	*36*	39	36	5
Washington	3	1,248	*15*	14	10	111
Westmoreland	17	461	*58*	68	68	55
SW Region total	111	3,331	397	431	409	303

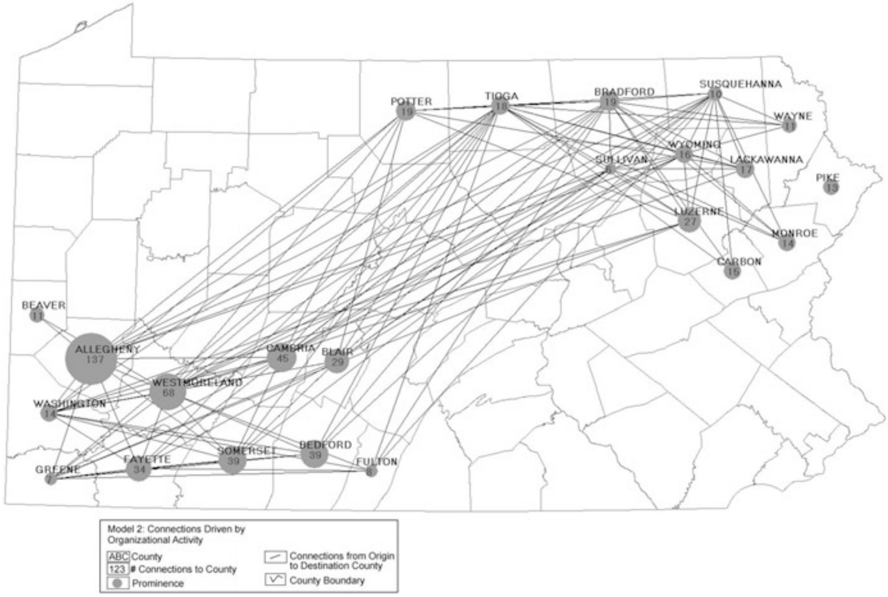

Fig. 11.7 Model 2: Origin constrained by connections in destination county

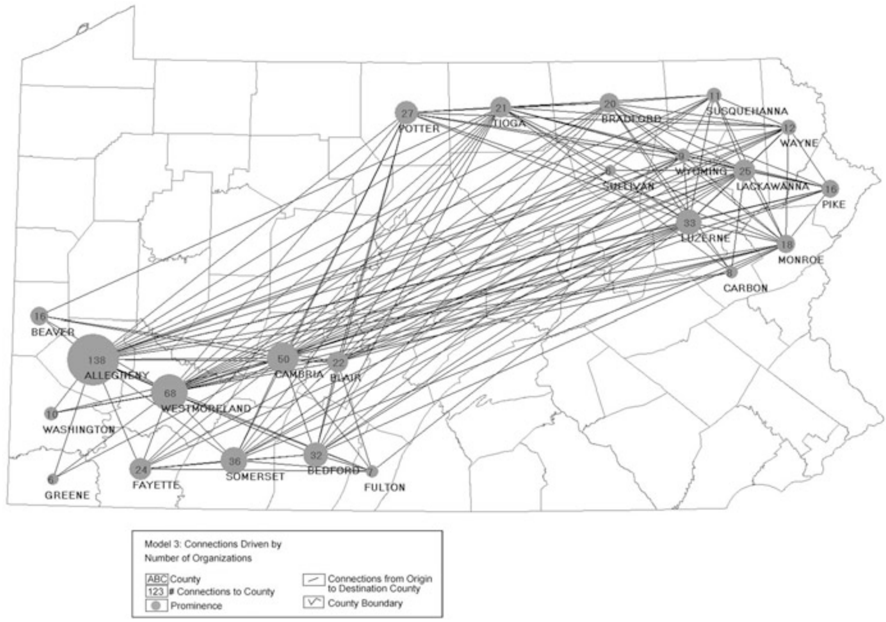

Fig. 11.8 Model 3: Origin constrained by number of organizations in destination county

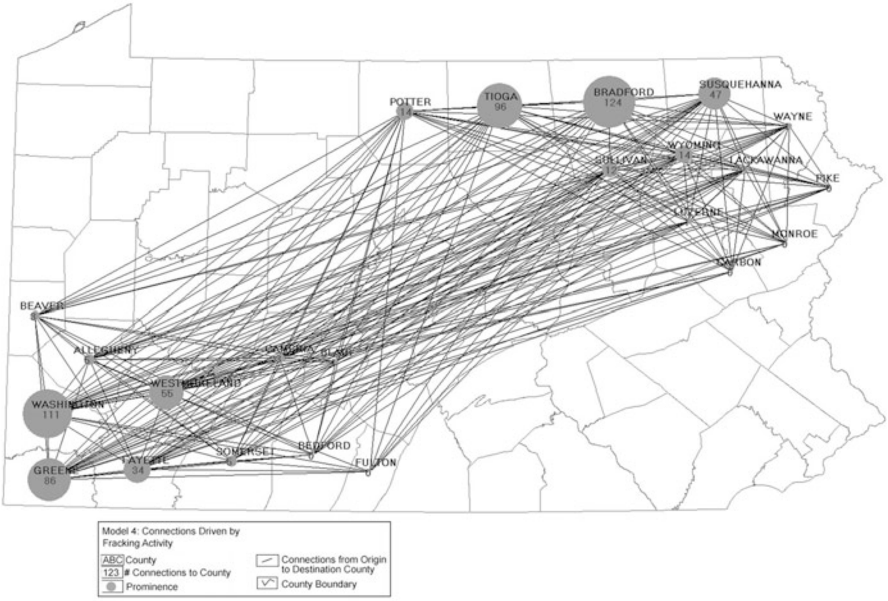

Fig. 11.9 Model 4: Origin constrained by fracking activity in destination county

> of and even control distant events. A similar point can be made about social space. . . Yet it is clear that some social spaces are closer together than others and that some social spaces are farther apart. While in the past, this was very much limited by geography, it can clearly be seen that geography, from a theoretical point of view, is really a stand-in for propinquity in social space. In the modern world, however, it is possible for fields that are not directly linked in geographic space to be socially connected. (McAdam and Fligstein 2012, p. 62)

In the case of the Marcellus environmental network, the juncture between social connections and geography remains relatively tight. In this case, geographic space can still be said to exert influences on social space. Where these spatial elements have forged organizational bonds across space, it is not unlikely that such patterns will build new social relations among these organizations.

Yet, while space and the spatial distribution of organizations do seem to shape the broad outlines of this network, this analysis identified several significant geographic holes in the advocacy network. Two of the least connected counties, Washington and Greene, are geographically proximate to two of the counties with the largest number of organizations and resources devoted to networking. Here at least the material influences of space seem absent and the reason for the dearth of connections sought in the social spaces of these communities. Indeed, the presence of an overall spatial pattern that fits this network landscape is not dis-confirmatory evidence of the influence of social spaces in this landscape. It does, however, suggest that there are complex interactions between material and social influences on network, collective action and community impacts. It is within this interactional complex that useful extensions of these findings could be sought. We will focus on three areas: (1) extensions for social movement approaches; (2) extensions for spatial sociology; and (3) implications for community problem solving.

We began by noting social movement's recent theoretical and substantive incorporation of community context into the analysis of these specific types of social interactions. McAdam and Boudet (2012) do so by making community characteristics a major analytic dimension of mobilization. For them, communities are the contextual arenas influencing success in mobilization. McAdam and Fligstein (2012) focus on communities as multiple interactional fields loosely positioned in space, but comprising "locality." By adopting community context as an analytic construct, these approaches necessarily confront many of the issues central to socio-spatial approaches: community composition, spatial limitations on interactions, spatial concentration and agglomeration, variations in normative embeddedness, and culture among places, etc. In confronting these issues the direction of social movement analysis begins to dovetail with theories, methods, and analysis in formerly tangential socio-spatial approaches. Sociology, demography, economics, and geography have well developed scholarly traditions that provide theoretical and substantive understandings of the relationships among space, place, and society. Melding socio-spatial scholarship into this line of social movement research can enhance both approaches.

Our analysis provides one example of this potential. Our analysis shifts focus from intra-community interactions to inter-community interactions, following a long line of socio-spatial analysis (Galpin 1915; Hawley 1950; Porter and Howell 2012, Chapter 2). By doing so, we have drawn attention to one influential element of context in mobilization the structure of spatial connections among individuals, organizations, and communities. In our study, this socio-spatial network grew from a relatively unconnected and ideologically diverse set of local community organizations into a geographically integrated complex. The spatial scope of this network clearly developed in response to the geographic scale of the environmental problems these communities confronted. The result is an expansion of one strategic action field that creates a wider community of action (McAdam and Fligstein 2012).

As our analysis has suggested, these connections were conduits for sharing resources, most importantly the sharing of information. Communities, their constituent organizations and populations, were advantaged or disadvantaged by their access to this network. By itself, this suggests that a community's position in a system of communities can be as important as other community contextual dimensions proposed by McAdam and Boudet (2012). It also suggests that inter-areal connections become more important when the scale and geographic scope of the environmental problems exceed the resources of any one community.

Finally, we have used several standard models of socio-spatial interaction to explore the spatial pattern of these connections. The number of advocacy organizations in a community was clearly the driving factor in creating linkages among communities. However, frictions of distance clearly moderated the likelihood those connections would be made between any pair of places. When there were greater distances between two communities, it was simply less likely that their organizations would make connections and share resources. Again, this suggests a spatial delineation of context in understanding impacts on and mobilization of communities. If the expansion of strategic action fields does follow network lines, then these fields will follow a size hierarchy (here defined by numbers of organizations) where communities of larger sizes predominate resource flows and become centers of mobilization activity. Because distance moderates interactions, these hierarchies are likely to regionalize. Communities of smaller sizes (fewer organizations) are less likely to develop a broad spatial scope of interaction. In this, the scope of the strategic action fields and the geography of contextual influence are likely to follow hierarchical settlement structures well known in sociology, geography and economics (Irwin and Kasarda 1994).

These observations demonstrate ways that socio-spatial scholarship might inform social movement studies. Certainly the central theoretical questions addressed by scholars of social movements are distinct from those posed by socio-spatial theory. However, socio-spatial approaches provide broad outlines, models of spatial interaction and social connections among communities that could contextualize the processes studied by social movement scholars.

It is also likely that the methodology and theoretical developments in social movement literature can expand these spatial models. Many of the standard models in socio-spatial analysis are deterministic extrapolations based on long held assumptions about individual and organizational interactions. They provide baseline comparisons that have theoretical value (Mayhew 1984), but are limited in their elucidation of the content and nature of ties and exchanges (Anderson 2011). This is especially true in examining the types of networks, such as discussed here, that link social and political individuals and organizations.

Systematic qualitative analysis of these data by Pischke suggested a relationship between propinquity, trust, and history (2013). One reason distance played an important role in determining connections lay in the character and regularity of personal connections. Such micro-level connections play out in routinized connections among organizations at the meso-level. Groups tend to collaborate when they have a history of doing so or when they are part of a larger umbrella organization (cf. Trout Unlimited affiliations). The trust in others to reciprocate resource sharing and funding is necessary for the groups to continue working together. Where trust and experience is greater, longer distances between groups may matter less. Similarly depersonalized contacts (cold-calling, written proposals, etc.) for assistance can place barriers to interaction regardless of propinquity.

Where these elements modify the effect of distance on interaction, holes will exist in the predictive models. Such network holes did exist in our analysis in two counties. Explaining these holes requires closer, systematic analysis of the history, motivations, interests and power differentials underlying these networks. Socio-spatial analysis seldom delves into such processes. However, the methodological approaches that mix qualitative and quantitative analysis suggested by McAdam and Boudet (2012), McAdam and Fligstein (2012), and surveyed in Diani and McAdam (2003) provide approaches that could well be adapted to this end. Developing analytic and conceptual linkages from micro to macro-social phenomenon through the melding of analytic approaches could well push scholars in socio-spatial and social movement research into new and potentially fruitful arenas.

A more complete picture of spatial processes and mobilization would also better serve individuals and organizations involved in community problems solving activity, such as those studied here. For the individuals and organizations involved in advocacy activity, understanding the larger structural implications of network position and understanding how they are linked to this network opens the door for more proactive involvement. Identifying holes in resource sharing networks is the first step to increasing connections to these underserved areas. Proactive network shaping can proceed through targeted outreach and programs.

In this specific case research results are of direct use for state and local policy initiatives dealing with resource mobilization surrounding Marcellus activities in Pennsylvanian communities (Brasier et al. 2011, p. 36; Jacquet 2009, p. 2). Support from environmental organizations that can lobby on behalf of local citizens and members will become necessary in a greater number of communities as hydro-fracturing spreads across the state. More generally, developing a holistic picture of the network can also address challenges in coordinating their advocacy efforts. This ensures that the geographic distribution of advocacy is not being duplicated and that limited resources are not wasted within the network. For these reasons, these middle range theoretical propositions generated by socio-spatial and social movement approaches can directly enhance advocacy impacts.

References

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics, 3*, 133–160.

Arthur, W. B. (1988). Self-reinforcing mechanisms in economics. In P. Anderson, K. Arrow, & D. Pines (Eds.), *The economy as an evolving complex system* (pp. 9–32). Reading: Addison-Wesley.

Baldassarri, D., & Diani, M. (2007). The integrative power of civic networks. *American Journal of Sociology, 113*(3), 735–780.

Belaire, J. A., Dribin, A. K., Johnston, D. P., Lynch, D. J., & Minor, E. S. (2011). Mapping stewardship networks in urban ecosystems. *Conservation Letters, 4*, 464–473.

Blau, J. R., & Rabrenovic, G. (1991). Interorganizational relations of nonprofit organizations: An exploratory study. *Sociological Forum, 6*(2), 327–347.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology, 92*, 1170–1192.

Brasier, K. J., Filteau, M. R., McLaughlin, D. K., Jacquet, J., Stedman, R., Kelsey, T. W., & Goetz, S. J. (2011). Residents' perceptions of community and environmental impacts from development of natural gas in the Marcellus Shale: A comparison of Pennsylvania and New York cases. *Journal of Rural Social Sciences, 26*(1), 32–61.

Bryan, T. (2004). Tragedy averted: the promise of collaboration. *Social and Natural Resources, 17*, 881–896.

Bullard, R. D., & Johnson, G. S. (2000). Environmental justice: Grassroots activism and its impact on public policy decision making. *Journal of Social Issues, 56*(3), 555–578.

Carmin, J. (1999). Voluntary associations, professional organizations, and the environmental movement in the United States. *Environmental Politics, 8*(1), 101–121.
Cathles, L. M. (2010). *The Marcellus gas resource*. Cornell University, Department of Earth and Atmospheric Sciences. Retrieved from http://cce.cornell.edu/EnergyClimateChange/NaturalGasDev/Documents/PDFs/The%20Marcellus%20Gas%20resource.pdf
Chief Oil and Gas. (2009). *The Marcellus Shale update*. Retrieved from http://www.chiefog.com/newsletters/0209/
Christopherson, S. (2011, September). *The economic consequences of Marcellus Shale gas extraction: Key issues* (CARDI reports 14). Ithaca, NY: Cornell University Community & Regional Development Institute.
Considine, T., Watson, R., Entler, R., & Sparks, J. (2009). *An emerging giant: Prospects and economic impacts of developing the Marcellus Shale natural gas play*. The Pennsylvania State University College of Earth and Mineral Sciences Department of Energy and Mineral Sciences. State College, PA: The Pennsylvania State University.
De Jong, T., & Van der Vaart, N. (2010). *Manual flowmap 7.4*. Utrecht University, The Netherlands. Retrieved from http://flowmap.geo.uu.nl/downloads/FM740_Manual.pdf
De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (2nd ed.). Cambridge: Cambridge University Press.
Diani, M & D. McAdam (2003). *Social movements and networks: Relational approaches to collective action*. Oxford: Oxford University Press.
Diani, M. (2003a). 'Leaders' or brokers? Positions and influence in social movement networks. In M. Diani & D. McAdam (Eds.), *Social movements and networks: Relational approaches to collective action* (pp. 105–122). Oxford: Oxford University Press.
Diani, M. (2003b). Introduction: Social movements, contentious actions, and social networks: From metaphor to substance. In M. Diani & D. McAdam (Eds.), *Social movements and networks: Relational approaches to collective action* (pp. 1–18). Oxford: Oxford University Press.
Environmental Resources Research Institute. (1998). *Pennsylvania – 104 state water plan watersheds, 1998*. University Park: Pennsylvania Gap Analysis Program. Environmental Resources Research Institute, The Pennsylvania State University.
Fisher, L. (2010). Allegheny conference: 70 percent of Marcellus Shale workers from out-of-state. As quoted in *PA Environmental Digest*. Retrieved from http://paenvironmentdaily.blogspot.com/2010/07/allegheny-conference-70-percent-of.html
Freudenburg, W. R., & Wilson, L. J. (2002). Mining the data: Analyzing the economic implications of mining for nonmetropolitan regions. *Sociological Inquiry, 72*(4), 549–575.
Galpin, C. J. (1915). The social anatomy of an agricultural community. *Research Bulletin* , p 34. University of Wisconsin Agricultural Experiment Station.
Glanville, J. L. (2004). Voluntary associations and social network structure: Why organization location and type are important. *Sociological Forum, 19*(3), 465–491.
Gould, R. V. (1991). Multiple networks and mobilization in the Paris commune, 1871. *American Sociological Review, 56*(6), 716–729.
Gould, R. V. (2003). Why do networks matter? Rationalist and structuralist interpretation. In M. Diani & D. McAdam (Eds.), *Social movements and networks: Relational approaches to collective action* (pp. 233–257). Oxford: Oxford University Press.
Hawley, A. H. (1986) *Human ecology: A theoretical essay*. Chicago: The University of Chicago Press.
Hawley, A. H. (1950). *Human ecology: A theory of community structure*. New York: The Ronald Press Company.
Headwaters Economics. (2009). *Fossil fuel extraction as a county economic development strategy: Are energy-focusing counties benefiting?* Retrieved from http://www.headwaterseconomics.org/energy/HeadwatersEconomics_EnergyFocusing.pdf
Hughes, H. L. (1993). Metropolitan structure and the suburban hierarchy. *American Sociological Review, 58*, 417–433.

Irwin, M. D. (2007). Territories of inequality. An essay on the measurement and analysis of inequality in grounded place settings. In L. Lobao, G. Hooks, & A. R. Tickamyer (Eds.), *The sociology of spatial inequality* (pp. 85–109). Albany: SUNY Press.

Irwin, M. D., & Hughes, H. L. (1992). Centrality and the structure of urban interaction: Measures, concepts and applications. *Social Forces, 71*(1), 17–51.

Irwin, M. D., & Kasarda, J. D. (1994). Trade, transportation and spatial distribution. In N. Smelser & R. Swedburg (Eds.), *The handbook of economic sociology* (pp. 342–367). Princeton/New York: Princeton University Press/Russell Sage Foundation.

Jacobson, M., & Kelsey, T. W. (2011). *Impacts of Marcellus shale development on municipal governments in Susquehanna and Washington counties, 2010*. Pennsylvania State University: *PennState Cooperative Extension Publications*. Retrieved from http://www.marcellus.psu.edu/resources/PDFs/jacobson_fiscal.pdf

Jacquet, J. (2009). *Energy boomtowns and natural gas: Implications for Marcellus Shale local community governments and rural communities*. The Northeast Regional Center for Rural Development. State College: The Pennsylvania State University.

Kargbo, D. M., Wilhelm, R. G., & Campbell, D. J. (2010). Natural gas plays in the Marcellus Shale: Challenges and potential opportunities. *Environmental Science & Technology, 44*(15), 5679–5684.

Kaufman, H. F. (1966). Toward an interactional conception of community. In R. L. Warren (Ed.), *Perspectives on the American Community* (pp. 88–103). Chicago: Rand McNally & Company.

Kerr, R. A. (2010). Natural gas from shale bursts onto the scene. *Science, 328*, 1624–1626.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*, 483–499.

Lake, R. L. D., & Huckfeldt, R. (1998). Social capital, social networks, and political participation. *Political Psychology, 19*(3), 567–585.

Lyson, T. A. (2004). *Civic agriculture: Reconnecting farm, food, and community*. Lebanon: University Press of New England.

Mayhew, B. H. (1984). Baseline models of sociological phenomena. *The Journal of Mathematical Sociology, 9*(4), 259–281.

McAdam, D., & Boudet, H. (2012). *Putting social movements in their place: Explaining opposition to energy projects in the United States, 2000–2005*. Cambridge: Cambridge University Press.

McAdam, D., & Fligstein, N. (2012). *A theory of field* (Kindle Edition). Oxford: Oxford University Press.

MIT. (2011). *The future of natural gas: An interdisciplinary MIT study. MIT energy initiative*. MIT University. Retrieved from http://mitei.mit.edu/publications/reports-studies/future-natural-gas

Murray, S., & Ooms, T. (2008). *The economic impact of Marcellus Shale in northeastern Pennsylvania*. Wilkes Barre: The Joint Urban Studies Center.

Pennsylvania Department of Environmental Protection (DEP). (2000–2012). *DEP oil & gas reporting website – Statewide data downloads by reporting period*. Retrieved from https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/DataExports/DataExports.aspx

Pennsylvania Department of Environmental Protection (DEP). (2010a). Marcellus Shale. Retrieved from http://www.elibrary.dep.state.pa.us/dsweb/Get/Document-77964/0100-FS-DEP4217.pdf. Retrieved 10/1/2010.

Pennsylvania Department of Environmental Protection (DEP). (2010b). *Active Marcellus Shale operators list*. Retrieved from http://www.dep.state.pa.us/dep/deputate/minres/oilgas/new_forms/marcellus/marcellus.htm. Retrieved 10/1/2010.

Pennsylvania Department of Environmental Protection (DEP). (2012). *Hydraulic fracturing overview*. Retrieved from http://files.dep.state.pa.us/OilGas/BOGM/BOGMPortalFiles/MarcellusShale/DEP%20Fracing%20overview.pdf

Pennsylvania Spatial Data Access (PASDA). (2013). *Pennsylvania – 104 state water plan watersheds, 1998*. Retrieved from http://www.pasda.psu.edu/uci/MetadataDisplay.aspx?entry=PASDA&file=majorsheds1041998.xml&dataset=15

Pischke, E. (2013). *Advocacy networks in the Marcellus Shale area: A study of environmental organizations in northeastern and southwestern Pennsylvania*. Unpublished master's thesis. Duquesne University, Center for Social and Public Policy, 186 pp.

Porter, J. R., & Howell, F. M. (2012). *Geographical sociology: Theoretical foundations and methodological applications in the sociology of location* (GeoJournal library, 105) (Kindle location, 686). Dordrecht: Kindle Edition.

Ravenstein, E. G. (1889). The laws of migration. *Journal of the Royal Statistical Society, 52*(2), 241–305.

Reeder, L. C. (2010). Creating a legal framework for regulation of natural gas extraction from the Marcellus Shale formation. *Social Science Research Network, 34*, 27.

Schmidt, C. W. (2013). Estimating wastewater impacts from fracking. *Environmental Health Perspectives, 121*, a117. Retrieved from http://dx.doi.org/10.1289/ehp.121-a117

Shankman, S. (2010, February 22). Gas drillers plead guilty to felony dumping violations. *ProPublica*. Retrieved from http://www.propublica.org/article/gas-drillers-plead-guilty-to-felony-dumping-violations

Snow, D. A., & Soule, S. A. (2010). *A primer on social movements*. New York: W.W. Norton & Company.

Soeder, D. J., & Kappel, W. M. (2009). *Water resources and natural gas production from the Marcellus Shale. USGS Fact Sheet 2009–3032*. Retrieved from http://pubs.usgs.gov/fs/2009/3032/

The Express. (2009, August 29). *Marcellus Shale topic of community meeting Tuesday: One-on-one time with industry reps and formal presentation planned*. Retrieved from http://www.lockhaven.com/page/content.detail/id/512687.html

The Pittsburgh Geological Society. (n.d.). *Natural gas migration problems in western Pennsylvania*. Retrieved from http://www.pittsburghgeologicalsociety.org/naturalgas.pdf

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(2), 234–240.

Tolbert, C. M., Lyson, T. A., & Irwin, M. D. (1998). Locality, civic engagement, and socioeconomic well-being. *Social Forces, 77*(2), 401–428.

US Census Bureau. (2007a). *Summary File 3. 2000 Census of population and housing technical documentation*. Washington, DC: Department of Commerce.

US Census Bureau. (2007b). *Summary File 1. 2000 Census of population and housing technical documentation*. Washington, DC: Department of Commerce.

US Department of Energy (DOE). (2009). *Modern Shale gas – A primer*. Retrieved from http://www.dep.state.pa.us/dep/deputate/minres/oilgas/US_Dept_Energy_Report_Shale_Gas_Primer_2009.pdf

US Department of Energy (DOE). (2010). Challenges facing developers of the Marcellus Shale in Appalachian Basin. *E&P Focus*, *1*(summer), 1–24. Retrieved from http://netl.doe.gov/technologies/oilgas/publications/newsletters/epfocus/EPNews2010Summer.pdf

US Environmental Protection Agency (EPA). (2013). *What is a watershed?* Retrieved from http://water.epa.gov/type/watersheds/whatis.cfm

Wilkinson, K. P. (1970). The community as a social field. *Social Forces, 48*(3), 311–322.

Zipf, G. K. (2012) [1949]. *Human behavior and the principle of least effort: An introduction to human ecology*. Mansfield Center: Martino Publishing.

Chapter 12
American Civic Community Over Space and Time

Charles M. Tolbert, F. Carson Mencken, Troy C. Blanchard, and Jing Li

In a number of papers we have elaborated extensively on what we mean by civic community. The perspective focuses on social and economic structures and institutions that buffer communities from external, usually global, forces. This leads us to identify and study important community organizations such as locally oriented business establishments, civic organizations, associations, churches, and an active electorate. These critical entities, in turn, are posited to benefit communities through an enhanced quality of life, more civic engagement by the citizenry, and a strong capacity for local problem solving (Tolbert et al. 1998, 2002; Blanchard et al. 2012).

The civic community perspective shares much with the declining social capital thesis typified by Putnam's *Bowling Alone* hypothesis. But, the work has been largely cross-sectional and unable to address fully themes like the decline of essential civic institutions or the hypothesis that social capital is in decline. Putnam (2000) does draw on theories of political regionalism that focus on broad areas of the U.S. Yet, there is an overriding implicit assumption that the decline in civil society is occurring evenly across the United States. Though some controls for regions of the U.S. have been employed, there has yet to be a detailed spatial analysis of variations in key measures of civic community. In this chapter, we address both temporal and spatial trends as we assess changes in levels of civic community.

This research was supported, in part, by grant no.s TEXR-2010-04719 and TEXR-2008-02636 from the National Institute of Food and Agriculture, U.S. Department of Agriculture.

C.M. Tolbert (✉) • F.C. Mencken • J. Li
Department of Sociology, Baylor University, Waco, TX 76798, USA
e-mail: Charlie_Tolbert@baylor.edu

T.C. Blanchard
Department of Sociology, Louisiana State University, Baton Rouge, LA 70803, USA

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_12

We develop an index of civic community based on a principal components factor analytic solution of American county data. The index is calculated at four points in time that correspond to the four most recent U.S. decennial population censuses: 1980, 1990, 2000, and 2010. We conduct analysis of spatial autocorrelation in U.S. counties at the four time points. We find substantial evidence of regional and local clustering of civic community across the U.S. We then use a pooled time-series, cross-sectional approach to model change in civic community over space and time. We find a decline in average levels of civic community between 1980 and 2010, but the trend is not linear. We also find that, net of significant spatial autocorrelation, several characteristics of counties facilitate the development of civic community structures over time (county levels of education, income, population size/density) while other characteristics impede the development of civic community structures over time (net migration, percent nonwhite, foreign born, and income inequality).

12.1 The Declining Social Capital Perspective

One of the most influential notions of recent social science is political scientist Robert Putnam's "Bowling Alone" thesis. He argues that the U.S. is suffering from a decline in civic engagement that threatens to undermine civil society. Putnam (1993) found a clear relationship between levels of civic engagement and socioeconomic development in his work in Italy during the 1990s. In the book *Bowling Alone*, Putnam (2000) changed his focus to the United States. Much evidence is presented about a decline in interaction and engagement among Americans. Putnam draws on the term *social capital* which he says " . . . refers to connections among individuals—social networks and the norms of reciprocity and trustworthiness that arise from them" (Putnam 2000:19). Americans have not just quit bowling together in organized leagues, Putnam contends, but they have disengaged from any number of organizations and associations. He points to negative consequences including less trust, less optimism about the future, less philanthropy, less volunteering, and so on. In a neo-Toquevillian way, rebuilding America's civic engagement becomes a path to meaningful citizenship, stronger democratic institutions, and an improved quality of life.

The declining social capital thesis has stimulated much interest, research, and controversy. In sociology circles, McPherson et al. (2006) published survey findings showing a decline in the number of close confidants along with an increase in socially isolated respondents (those with reporting no confidants). Though widely disseminated through the media, the findings were later criticized by Fischer (2009), among others, who argued that the results were an artifact of the survey question construction. Nonetheless, the declining social capital thesis remains one of the most provocative of early twenty-first century social science. However, our civic community work demonstrates that the structural features of place affect the development of social capital. Incorporating concepts from civic community perspectives will help us understand why the decline in social capital is uneven across both time and space.

12.2 The Civic Community Perspective

Among social scientists who study rural areas, an institutional perspective has flourished alongside the broader interest in social capital. The focus is on community social structures (small business, associations, organizations) that encourage civic engagement. It is well established that locally-oriented businesses such as small manufacturing establishments and retail outlets are associated with a number of beneficial local outcomes and promote community resilience (e.g., Tolbert et al. 2002; Blanchard et al. 2012). The benefits include better socioeconomic conditions (e.g., higher income levels, less poverty, less income inequality, lower unemployment), lower crime and delinquency levels, and even better public health outcomes (Tolbert 2005; Blanchard and Matthews 2006). This line of work fits in a broader social science rural development research agenda that calls attention to social structure and problem-solving capacities of an engaged citizenry. Writers have used terms such as corporate social responsibility, entrepreneurial infrastructure, local capitalism, civic community, and social networks (Besser 1998; Flora et al. 1992, 1997; Green 2003; Tolbert 2005; Blanchard and Matthews 2006). Implicit in this work is the notion that building a community's stock of civic capital will yield positive outcomes very similar to those to be had by promoting social capital.

We argue that the research on civic community, with its focus on the impact of local structures such as small businesses, third places, and associations (Tolbert et al. 1998, 2002; Blanchard et al. 2012), needs better integration with the broader theoretical macro-level systemic model of community structure and organization. Accordingly, systems of local kinship and friendship networks intersect with formal and informal associational ties to create a greater sense of community identity and integrate individuals into active roles in community life (Kasarda and Janowitz 1974; Sampson 1988; Sampson and Groves 1989). While locally based social networks are at the heart of systemic community formation, at a conceptual level three structural features of places (population stability, population homogeneity, socioeconomic resources) are exogenous to the formation of social networks and collective identity.

Proponents of the systemic model (Kasarda and Janowitz 1974) emphasize the importance of residential stability for community structure. Locales with high levels of population turnover have low levels of social and kinship network formation. It is difficult to establish effective relationships in a population that is frequently churning. Second, population homogeneity affects bridging network formation. Social networks have a difficult time reaching across structural barriers of class, race, and ethnicity (McPherson et al. 2001; DiPrete et al. 2011; DiMaggio and Garip 2011). Therefore, communities with more heterogeneous populations will have less bridging capital and lower levels of civic community and social capital. These two concepts have not been included in our previous models of civil society.

In the past we have modeled socioeconomic disadvantage as an endogenous variable. However, it can affect civic community through creating a milieu of apathy and dampened community attachment. This has a negative effect on civic activities, such as volunteering and voting (Davenport 2010; Sondheim and Green 2010). Lack of involvement, in turn, can lead to low levels of community attachment and network formation. Socioeconomic disadvantage can also lead to low levels of trust in neighbors, especially in places with high levels of social problems, such as drug use and violent crime (Bellair 1997; Sampson and Byron Groves 1989). In addition, residents who are socioeconomically disadvantaged may not possess the cultural capital necessary to participate comfortably in local affairs. This has the potential to be particularly problematic for new immigrant populations who may not speak the language well (Perlmann 2005; Terriquez 2012).

One of the methodological strengths of this line of work is that organizations lend themselves to measurement at the community (usually county) level. This has permitted researchers to employ quantitative spatial econometric approaches to the analysis of civic community (Tolbert et al. 1998; Blanchard et al. 2012). Yet, longitudinal research designs that might test for change in the level of civic community have been scarce (Tolbert 2005). In this paper, we introduce civic community response measures that vary by both space and time. Based on our theoretical development, we include three conceptual predictors of civic community: population stability, population homogeneity, and socioeconomic resources. We also control for other demographic factors, including population size and density. This allows us to test for declining levels of civic community with a theoretically informed analysis while controlling for spatial unevenness in the distribution of civic community.

12.3 Data and Methods

We present two stages of analysis. First, we perform a LISA (local indicators of spatial autocorrelation) analysis of the spatial and temporal distribution of civic community among U.S. counties for 1980, 1990, 2000 and 2010. Second, to evaluate the covariates of change in civic community over time, we estimate a fixed effect regression model for panel data with a spatial lag term using PROC GLM in SAS (see Allison 2005). Data on our 3,059 counties are structured as county years with four observations for each county (1980, 1990, 2000, and 2010), yielding a total of 12,236 observations. Analysis of panel data is more complex than cross-sectional because variation in the dependent variable can be parsed into to changes over time within a county and difference in baseline levels of the dependent variable between counties.

12.3.1 Dependent Variable

We draw from previous work and use five measures of civic community to construct a *civic community index*. Three of the civic community items are based on tabulations of businesses and organizations in *County Business Patterns*: associations per 10,000 residents, small manufacturing establishments per 10,000 residents, and third places per 10,000 residents (Tolbert et al. 1998; Irwin et al. 2004). Earlier work has shown several beneficial community outcomes associated with these local business indicators (see Blanchard et al. 2012; Tolbert et al. 1998). Two non-business civic measures are also part of the principal components solution. The percent of persons age 18 and older voting in the most recent presidential election (1980, 1988, 2000, and 2008) is taken from *USA Counties*. The number of adherents in civically engaged denominations per 10,000 residents is derived from the decennial denomination studies by the Association of Statisticians of American Religious Bodies and has been used in our past research on regional growth (Mencken et al. 2006). Data for some official county equivalents such as the Virginia independent cities are aggregated into the data values for the surrounding counties. The county boundaries are harmonized over time (since 1980) to ensure consistency in spatial units. In these ways, we arrive at a time-space dataset with uniform temporal and spatial attributes.

We measure civic community using a factor score developed from a principal components analysis of the five indicators of civic community described above. Given the longitudinal nature of our data, we construct both a within decade and across decade factor analysis. In the within decade factor analysis, the unit of observation is the county and we develop four factor scores, one for each decade for the 3,059 counties under study. The within decade factor score provide a means to observe how the spatial patterning of civic community may change over time. We examine these decade specific factor scores using a LISA analysis. Our second factor score is calculated across the entire time series. The unit of observation in this factor analysis is the county-year (3,059 counties measured at four point in time yielding 12,236 county-years). In this factor score, we are able to identify the level of civic community at different points to measure the change in our civic community factor score. We use this factor score in our multivariate analysis to understand how change in demographic, social, and economic characteristics of counties is associated with changes in civic community.

In Table 12.1, we report the factor loadings from our principal components analysis. All the solutions result in one civic community factor. The loadings are at least 0.5 with the third places measure consistently showing the strongest relationship with the underlying factor. We produced factors scores for each solution. For the decade specific factor scores, a value of zero indicates an average level of civic community for that decade. In our factor analysis that encompasses all four decades, a score of zero indicates an average level of civic community across the entire four decade time series.

Table 12.1 Factor loading for civic community index

		Decade specific factor scores			
	Across time series	1980	1990	2000	2010
Small manufacturing per 10,000 residents	0.568	0.503	0.384	0.605	0.65
Third places per 10,000 residents	0.763	0.668	0.795	0.809	0.808
Associations per 10,000 residents	0.719	0.715	0.725	0.707	0.69
Adherents to civic denominations per 10,000 residents	0.414	0.551	0.557	0.555	0.514
Number voting in most recent presidential election per 10,000 residents age 18 and over	0.600	0.381	0.605	0.613	0.585
Eigenvalue	1.954	1.661	1.982	2.205	2.159

Independent Variables Our primary concern is to understand how changes in the dependent variable over time are associated with changes in the independent variables. Variation within a county over time is addressed through the inclusion of time varying independent variables. For example, decennial counts of population size or median family income for each county would be considered time varying independent variables.

We also include a special type of time varying independent variable to control spatial clustering in the dependent variable within each decade. Using GeoDa 1.4.0, we calculated the spatial lag in the dependent variable for each decade using a queen weight matrix with first order contiguity. The spatial lag for each county in each decade was then merged with our county years file. In our models, the spatial lag term is a time varying independent variable.

Our model also includes fixed effect terms that control for all time invariant characteristics of a county, also referred to as unobserved heterogeneity. Fixed effects are measured by including binary variables for each county. Although the fixed effect terms do not specify the source of the unobserved heterogeneity, the binary variables account for all between-county variation in the dependent variable. In doing so, model estimates for the time varying independent variables are not biased by unobserved heterogeneity among the units of analysis.

Our use of a fixed effect model for panel data that integrates a spatial lag term is based on recent research on the analysis of spatial panel data (for a review, see Elhorst 2010). The use of fixed effects alone in cross sectional spatial analyses has been criticized as inferior to spatial lag or spatial error models (Anselin and Arribas-Bel 2013). Within the context of panel data, fixed effect models are routinely used to account for temporal and observation level effects, such as survey respondents or ecological units (Allison 2005). When extended to spatial panel data, researchers employ a mix of fixed effects and corrections for spatial dependence—such as spatial lag terms and/or spatial error models—as we do below.

Descriptive statistics for the variables in our analysis are provided in Tables 12.2 and 12.3. In Table 12.2, we report means and standard deviations of the independent variables for each decade. We include three sets of predictors from the

Table 12.2 Descriptive statistics for the control variables

	1980		1990		2000		2010	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Population density (1,000 per square mile)	0.223	2.774	0.225	2.607	0.249	3.025	0.265	3.121
Metro	0.203	0.402	0.231	0.422	0.263	0.441	0.345	0.475
Log population size	10.112	1.316	10.141	1.370	10.237	1.407	10.281	1.459
Net migration	10.149	20.697	0.676	15.043	7.534	13.756	4.668	11.220
Percent Hispanic	3.793	10.251	4.397	11.094	6.151	12.144	8.309	13.276
Percent black	8.548	14.390	8.529	14.303	8.706	14.454	8.828	14.432
Percent foreign born	2.018	2.721	2.171	3.483	3.377	4.696	3.987	4.980
Gini coefficient	0.368	0.036	0.380	0.038	0.434	0.037	0.431	0.037
Median family income ($1,000 s)	16.694	3.506	28.251	6.920	41.939	9.690	52.967	12.626
Percent Hs grad	34.652	7.628	44.479	7.358	50.803	6.818	53.930	7.224
Spatial lag	0.815	0.789	−0.300	0.643	0.310	0.652	−0.237	0.618

Table 12.3 Descriptive statistics for the across decade civic community factor score

Year	Mean	SD
1980	0.869	206.511
1990	−0.457	148.853
2000	−0.430	161.891
2010	−0.357	165.928

All means significantly different from other decades except 1990 and 2000 using Tukey HSD $p < .05$

systemic model of community formation. Three measures of county population homogeneity are used: the percentage of population black, the percentage of Hispanics, and the percentage of foreign-born population. Three measures of socioeconomic resources are utilized, these are median family income (in thousand dollars), the gini coefficient, and the percentage of persons 25 years and over with at least a high school education. One measure of population stability is included in the model: the estimated county net migration rate for the previous decade. Two variables are included as controls for size of place- the natural logarithm of the county population size and population density (1,000 persons per square mile). Metropolitan status is a binary control variable based on the Rural-Urban Continuum Codes developed by United States Department of Agriculture (USDA) Economic Research Services. It is coded 1 if the county is in a metro area and 0 if it is not.

12.4 LISA Analysis of Civic Community

Using GeoDa, we conduct local indicators of spatial autocorrelation (LISA) analysis on the civic community factor index. We do this for four cross-sections of data that correspond to four recent decennial Census time points: 1980, 1990, 2000, and 2010. The LISA analysis indicates the extent of statistically significant spatial clustering of similar index values. Minimal clustering is indicative of a process that is spatially independent. In this case, it would mean that values on the civic community index are distributed relatively evenly across U.S. counties. Substantial clustering suggests the opposite; i.e., there are clusters of high levels of civic community in some parts of the country and clusters of low levels of civic community in other places. If an uneven distribution of index values is observed, then it is plausible that civic community may be strong in some areas and weak in others. This would cast doubt on assertions of a global decline in American civic community.

Figure 12.1a–d show LISA results for the four decennial cross-sections. Regional clusters of high civic community are indicated by the red ("hot spots") on the color versions of the maps and as black on the grayscale maps. Regional clusters of low civic community ("cold spots") are shown in blue or medium gray. We also report Moran's I, which is a measure of the extent of spatial clustering. A high Moran's I value indicates significant regional clustering of civic community values. Figure 12.1a presents the distribution and clustering of civic community for 1980. The map shows significant regional clustering of counties with high levels of civic community in the northeast, mid-Atlantic region, Midwest and Great Plains, as well as the Rocky Mountains and Pacific Northwest. In contrast, regional clusters of low values of civic community cluster are primarily in the South. These cold spots correspond to an area of persistent poverty often referred to as the Black Belt (Wimberley and Morris 2002; Allen-Smith et al. 2000). Other southern cold spots can be seen in historical high poverty concentrations (see Isserman et al. 2009) along the Mississippi delta (black poverty), in central Appalachia (white poverty), and along the Texas borderlands (Hispanic poverty). The Moran's I is 0.518 which indicates moderate clustering.

The Moran's I for the 1990 cross-section depicted in Fig. 12.1b is 0.64. This means that considerably more clustering of like counties has taken place across one decade. High civic community hot spots in 1990 are concentrated in the Great Plains region and the Rocky Mountains. The clustering of hot spots in the Great Plains and upper Mid-West- in particular the Upper Peninsula of Michigan- is puzzling, given the population outmigration from these regions during the 1980s

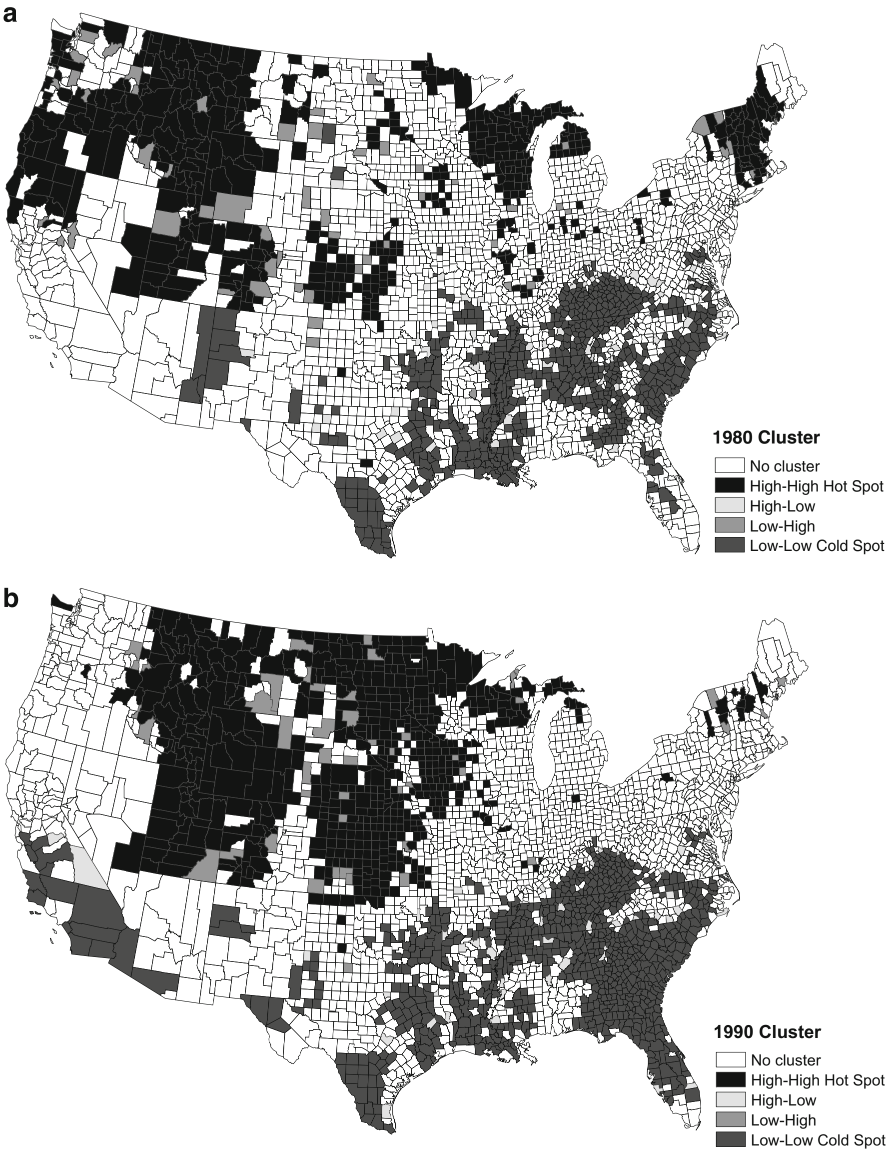

Fig. 12.1 (**a**) Clustering of Civic Community: 1980 (Moran's I = 0.518). (**b**) Clustering of Civic Community: 1990 (Moran's I = 0.644)

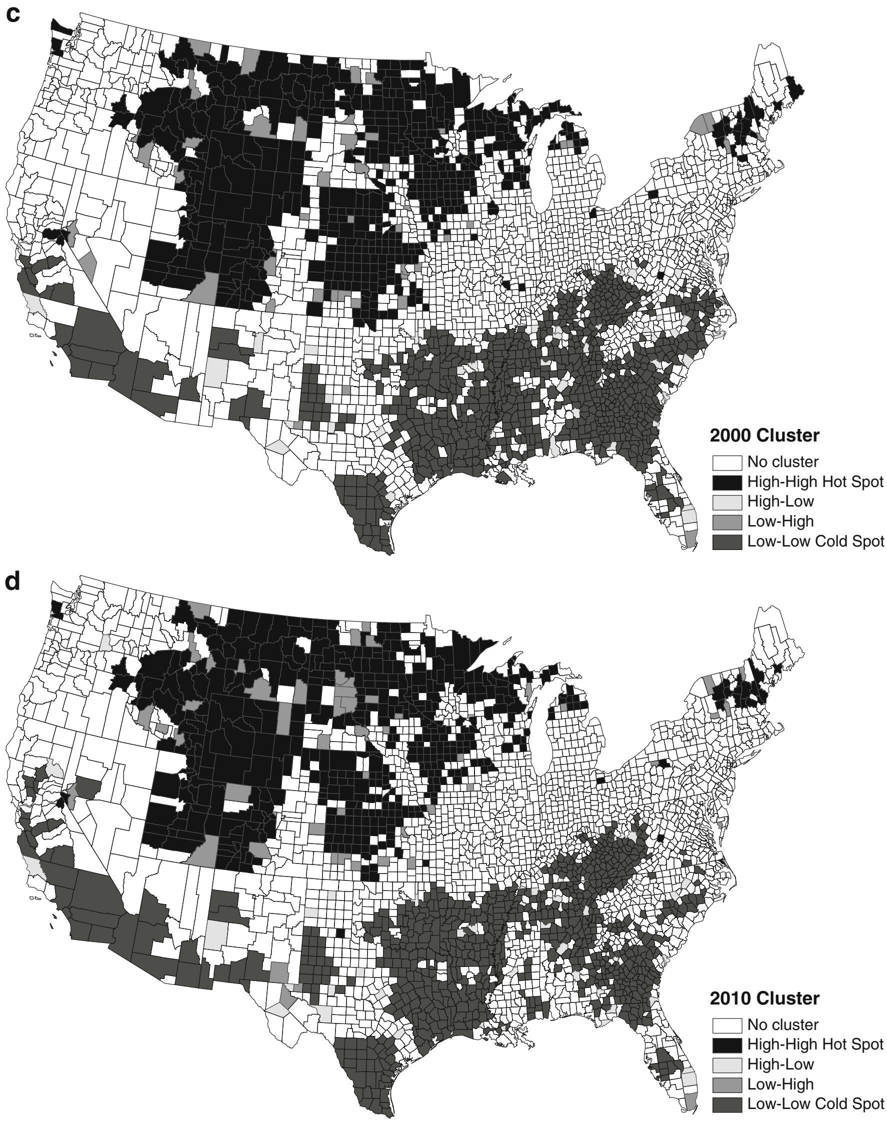

Fig. 12.1 (continued) (**c**) Clustering of Civic Community: 2000 (Moran's I = 0.576). (**d**) Clustering of Civic Community: 2010 (Moran's I = 0.517)

(see Schwarzweller and Lean 1993; Johnson and Rathge 2006; Rathge and Highman 1998). One plausible explanation is the outmigration stream of young people who were less inclined to be members of traditional civic organizations (Putnam 2000). The organizations still persist, but likely show a sharpening upward shift in age composition.

The clustering of hot spots in the Rocky Mountain region may reflect several changes. One is the potential emergence of local natural amenities as drivers of regional economic development. In a post-industrial professional service oriented economy the work place has greater geographical elasticity. The quality of life for professional workers becomes an important variable in deciding where to locate operations facilities. Amenity rich regions such as the Pacific Northwest and Rocky Mountains become popular destinations for new economy enterprises (McGranahan 1999, 2008). Tourist counties in Colorado (San Juan, Ouray) have the highest civic community scores of all US counties across the 1980–2010 time line. A second factor is the growth of civically engaged denominations, in particular the Latter Day Saints church throughout parts of Utah, Idaho, and adjacent counties in Montana and Wyoming (Cragun and Phillips 2008). The growth of a civically engaged denomination would help explain the development of hot spot clusters of civic community throughout this region of the U.S.

The Moran's I measures for 2000 and 2010 remain above 0.5 though there is a decline in observed clustering (0.58 for 2000 down to 0.52 for 2010). Figure 12.1c and d are similar in that they show most of the southern half of the U.S. as cold spots. Many of these cold spots are located in southern regions of historically high levels of poverty, such as the Black Belt and Lower Mississippi Delta regions. High levels of poverty lead to low voter turnout, low levels of socioeconomic resources, fewer civic organizations, and less capital to develop local businesses (Stoll 2001; Tomaskovic-Devey and Roscigno 1997). The second trend these maps document is the emergence of cold spots along the southern U.S. border. This pattern likely reflects the in-migration of Hispanics from Mexico and Central America. Perlmann (2005) finds that Hispanics, much like Italians who immigrated to the U.S. a century earlier, are not particularly civically engaged. As an outsider group that has been demonized by some, they may feel less welcomed and be less trusting of others. Lacking cultural capital, such as strong language skills, they are less likely to volunteer or be involved with civic organizations (Terriquez 2012). Conversely, the influx of new immigrants will diminish the willingness of the native population to engage their communities (Bell 2009; Coffe' 2009; Jobes 1999). An influx of new faces can create a 'bunker mentality' among the native population. New faces, especially those culturally different, can represent a threat to native populations, thus diminishing their willingness to engage in their communities (Coffe' 2009).

Clearly, change in American civic community has not occurred evenly over space and time. The Fig. 12.1a–d maps show a consistent north-south divide such that civic community hot spots are generally in the northern U.S. and colds spots are mostly in the South. Already poised for rapid social change after the civil rights era, the South experienced several waves of in-migration across the periods that we observe. These include the north to south "Sunbelt" migration, the return migration of African-Americans from the north, and the arrival of immigrants from south of the border. However, it is clear that the cross sectional maps I (Fig. 12.1a–d) raise as many questions about the temporal and spatial distribution of civic engagement as they answer. In order to develop a clearer understanding of the social forces that drive the spatial and temporal distributions of civic engagement we turn now to a multivariate model.

12.5 Cross-Sectional Time Series Fixed Effects Model

The map analysis in Fig. 12.1a–d provides insight into the changing spatial clustering of civic community, but tells us little about changes in the absolute level over time. To identify changes in the level of civic community over time, we report statistics from our across decade factor score in Table 12.3. The findings indicate that the average county in 1980 had a level of civic community that was higher than average counties in later decades. Furthermore, the average level of civic community stabilized between 1990 and 2000 and then significantly declined again from 2000 to 2010. The measures in Table 12.3 indicate that the average overall level of civic community has declined overall from 1980 to 2010. But, our LISA findings suggest this has occurred in an uneven fashion. This is important because prior research implicitly suggests a uniform secular decline in the level of civic community.

We build on our univariate analysis by estimating a fixed effect longitudinal regression model to examine the covariates of change in levels of civic community. We report the results of our fixed effect panel analysis in Table 12.4. In Model 1, we include our binary variables for each decade with 1980 as the contrast group. Although not shown in the table, our model also includes 3,058 binary variables to control for between county differences in the average level of civic community. Combined, these variables account for 90.9 % of the variation in the dependent variable. When disaggregated into between and within county variation, the county binary control variables capturing the time invariant differences between counties account for 49.6 % of the variation in the dependent variable. Time varying differences within counties in civic community represent 50.4 % of the variation. Within the context of the fixed effect analysis, the general trends in levels of civic community observed in Table 12.2 are in Model 1. Civic community levels in 1980 are significantly higher than in other decades. In addition, levels in 1990 and 2010 appear to be lower than the average level of civic community in 2000.

In Model 2, we add our spatial lag variable. The coefficient for the spatial lag variable is positive and significant indicating that the dependent variable exhibits positive spatial autocorrelation (e.g. counties with increasing levels of civic community are more likely to be surrounded by counties with increasing levels of civic community). Put differently, changes in civic community for a county are, in part, dependent on changes in adjacent counties. Thus, not only are levels of civic community spatially correlated at a cross-sectional level, temporal changes in

Table 12.4 Fixed effect regression model with spatial lag term

	Model 1			Model 2			Model 3		
	Coefficient	St. Err		Coefficient	St. Err		Coefficient	St. Err	
Intercept	0.465	0.161	**	0.247	0.141		−2.283	0.282	***
t80 (contrast group)	–	–		–	–		–	–	
t90	−1.301	0.007	***	−0.551	0.015	***	−0.648	0.022	***
t00	−1.256	0.007	***	−0.531	0.015	***	−0.615	0.034	***
t10	−1.164	0.007	***	−0.488	0.014	***	−0.629	0.041	***
Spatial lag	–	–		0.616	0.012	***	0.611	0.011	***
Population density	–	–		–	–		0.016	0.001	***
Metro	–	–		–	–		0.016	0.014	
Log population size	–	–		–	–		0.267	0.019	***
Net migration	–	–		–	–		−0.005	0.000	***
Percent Hispanic	–	–		–	–		−0.016	0.001	***
Percent black	–	–		–	–		−0.003	0.001	***
Percent foreign born	–	–		–	–		−0.010	0.002	***
Gini coefficient	–	–		–	–		−1.329	0.179	***
Median family income	–	–		–	–		0.007	0.000	***
Percent Hs Grad	–	–		–	–		0.003	0.001	*
R2	0.909			0.931			0.944		

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$ (two-tailed test)

civic community are similarly correlated. The substantial level of spatial autocorrelation in the analysis also indicates that the scope of civic structures spills over well beyond the boundaries of single counties.

The coefficients reported in the last column of Table 12.3 provide insight into covariate explanations of the changing U.S. civic community landscape. The key findings show that all three conceptual dimensions of the systemic model have important net effects on civic community. Our measure of population stability, estimated net migration rate for the previous decade, is associated with a decline in the level of civic community. Places with high levels of net immigrants have lower levels of civic community, and this relationship is consistent over time. This finding is consistent with the predictions of the systemic model. The in-migration of new faces leads to a breakdown in social networks. In some circumstances, new faces can represent a 'threat' to native residents, leading to lower levels of trust and a diminished willingness to continue to stay engaged in the community (Coffe' 2009).

Our findings from Model 3 also indicate that increasing racial heterogeneity is associated with declines in civic community. Changes in both the percent black and the percent Hispanic are associated with decreases in the level of civic community. These findings are consistent with past research which concludes that racial and/or ethnic population heterogeneity is associated with problems in developing bridging social ties within communities (McPherson et al. 2001; DiPrete et al. 2011; DiMaggio and Garip 2011). Moreover, Hispanics and African Americans have traditionally lower levels of key civic engagement, particularly voting and volunteering (Stoll 2001; Green and Gerber 2008; Musick et al. 2000; Wilson 2000; Einolf 2009). A greater concentration of these population groups leads to lower levels of civic community, *ceteris paribus*.

The percent foreign born in the county population has a significant effect on levels of civic community. The literature on population homogeneity provides a number of reasons why new residents may disrupt civic community, including lack of cultural capital (language, norms/customs) and barriers to network formation. It can also be argued that foreign born populations represent the worst cultural 'fit.' They are the least likely to know the native language and customs (Perlmann 2005; Terriquez 2012).

The measures for socioeconomic resources behave as predicted. The coefficients for median family income and educational attainment indicate that increases in socioeconomic status in the population are associated with increasing levels of civic community. Between 1980 and 2010, an increase in median family income yielded increases in civic community. The same finding holds for the percent of the population age 25 and older with at least a high school education. In contrast, the gini coefficient, which measures the relative distribution of income (income inequality), is associated with lower levels of civic community. Communities with higher poverty rates and unequal income distributions as well as lower median family incomes are resource disadvantaged communities that struggle to build the infrastructure necessary to facilitate the development of civically engaged citizenry.

The coefficients for the control variables show that increases in population density and population size are associated with increases in the level of civic community. Put differently, losing population or becoming more sparsely populated is associated with declines in civic community. The spatial lag variable is statistically significant, indicating that the statistically significant clustering of hot and cold spots presented in the Fig. 12.1a–d can partially, but not entirely, be explained by our set of predictors.

12.6 Discussion and Conclusions

This chapter has two purposes: (1) to contribute to the overall theme of this book by theorizing a spatialized conception of trends in civic community; and (2) to provide a sub-national test of Putnam's thesis that social capital has declined over time. Regarding the second objective, our analysis does provide some support for Putnam's claim. In our time frame, civic community peaked in 1980, and has been on a downward sloping roller coaster ride ever since. However, we do observe widespread spatial variation in civic community by 2010, indicating that civic institutions do not exhibit a monotonic, uniform decline across America. This suggests that claims of sweeping national changes in civic institutions amount to a gross overgeneralization. Our analysis shows that there are indeed pockets of civic decline that correspond to the thesis of national secular decline. Our analysis also shows that there are areas where civic community is sustained and even flourishing. In sum, by bringing a conventional analysis over space and time to a current issue, we demonstrate the plausibility and promise of a spatialized approach.

This chapter has also served to improve our basic research framework. In the past we have modeled two dimensions of local structure as exogenous predictors of civic welfare: local capitalism and civic engagement institutions (see Tolbert et al. 2002). Unlike our previous research, our analysis here conceptualizes civic engagement institutions as an endogenous dimension. Measures of systemic dimensions (population stability, population heterogeneity, existing resources) have been largely absent in our previous attempts. Given the expected continuation of spatial variation in future migration and fertility, our understanding of place well-being going forward will be enhanced by models that continue to integrate other important social processes (such as net migration, foreign born population composition).

On a final note, this chapter will help to stimulate further research on spatial inequality using sub-national analysis. Historically, much of the research on spatial processes has either been done with nations as the units of analysis, or by focusing exclusively on major urban areas. And, almost all of the research is cross-sectional. Our analysis examines and documents regional concentrations of civic community structures that vary over space and time. The longitudinal methods we employ allow us to say with statistical security that the regional clustering of civic community 'hot spots' and 'cold spots' signals important regional processes of uneven development. The importance of the spatial lag variable in the regression models indicates that this clustering is not fully explained by traditional ecological variables such as population density and size. Research on spatial inequality has done a good job of informing the literature on why places and regions are performing poorly (see Lobao et al. 2007). We maintain that our agenda on the sub-national spatial and temporal distribution of civic community will help to explain why some communities are doing well in the global economy. We hope these findings aid policymakers as they strive to eradicate the inequalities of past uneven development.

Regarding the first objective of this paper, to develop mid-range theory about the spatialization of civil society, our results do point to potential hypotheses for future research. In this paper, we test empirically whether there is unevenness in the spatial distribution of civic community. Our findings have some obvious implications for residents of areas that are relatively high or low on our civic institution metrics. For some, there will be increased chances of building social capital through engagement of others in associations, gathering places, and churches. In those same areas, a larger proportion of small manufacturing establishments suggests a more cooperative business climate that can foster firm startups, expansions, and sustain the local economy. High voter turnout is suggestive of a collective efficacy that promotes local problem-solving. For those in low civic areas, there will be fewer such opportunities to build social capital, less cooperative among local businesses, and dampened collective efficacy. Communities will benefit in the former case, but not so much in the latter.

Beyond the impacts on individuals, this is a potentially important set of results for theories of rural development and for development methodologies. The findings could lead to improved specification of the base civic community model in terms of contexts where the thesis is most likely to hold. Strong evidence of clustering of areas similar on civic community (high or low levels) would suggest that rural development policies need to take into account the levels of civic community in the immediate *and adjacent areas*. It may be possible, for example, to leverage the influence of a strongly civic place to have a beneficial spillover effect on neighboring communities. One example which informs this proposition is Fayette County, Georgia. In 1970, there were slightly over 11,000 residents living in the county. In 1980 it ranked in the bottom 25th percentile on the civic engagement index. By 2010 Fayette County had over 110,000 residents, and ranked in the top 25th on the civic engagement measure. Fayette County is in the greater Atlanta metropolitan area. The adjacency of this county to a thriving metropolitan area over time promoted the development of high levels of civic engagement infrastructure. While previous research on spatial economies underscores the importance of adjacency for regional economic growth, our findings indicate that the same processes may help to explain the development of regional centers of civic engagement. By the same token, development in a low civic community area may be hampered by an absence of nearby places with stronger civic climates. If we do not observe clustering by levels of civic community, then development practitioners can safely encourage the development of civic community institutions almost anywhere and reasonably expect beneficial outcomes.

Bibliography

Allen-Smith, J. E., Wimberley, R. C., & Morris, L. V. (2000). America's forgotten people and places: Ending the legacy of poverty in the rural south. *Journal of Agricultural and Applied Economics, 32*, 319–329.

Allison, P. (2005). *Fixed effect regression methods for longitudinal data in SAS*. Cary: SAS.

Anselin, L., & Arribas-Bel, D. (2013, March). Spatial fixed effects and spatial dependence in a single crosssection. *Papers in Regional Science, 92*(1), 3–17.

Bell, S. E. (2009). 'There Ain't No Bond in Town Like There Used to Be': The destruction of social capital in the West Virginia coalfields. *Sociological Forum, 24*, 631–657.

Bellair, P. E. (1997). Social interaction and community crime: Examining the importance of neighborhood networks. *Criminology, 35*, 677–704.

Besser, T. L. (1998). The significance of community to business responsibility. *Rural Sociology, 63*, 412–431.

Blanchard, T., & Matthews, T. L. (2006). The configuration of local economic power and civic participation in the global economy. *Social Forces, 84*, 2241–2256.

Blanchard, T., Tolbert, C. M., & Carson Mencken, F. (2012). The health and wealth of U.S. counties: How the small business environment impacts alternative measures of development. *Cambridge Journal of Regions, Economy, and Society, 5*, 149–162.

Coffe', H. (2009). Social capital and community heterogeneity. *Social Indicators Research, 91*, 155–170.

Cragun, R., & Phillips, R. (2008). *Mormons in the United States 1990–2008: Socio-demographic trends and regional differences.* A report based on the American Religious Identification Survey. Hartford: Trinity College.

Davenport, T. (2010). Public accountability and political participation: Effects of a face-to-face feedback intervention on voter turnout of public housing residents. *Political Behavior, 32*, 337–368.

DiMaggio, P., & Garip, F. (2011). How network externalities can exacerbate intergroup inequality. *American Journal of Sociology, 116*, 1887–1933.

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology, 116*, 1234–1283.

Einolf, C. J. (2009). Will the boomers volunteer during retirement? Comparing the baby boom, silent, and long civic cohorts. *Nonprofit and Voluntary Sector Quarterly, 38*, 181–199.

Elhorst, J. P. (2010). Spatial panel data models. In M. M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 377–407). Berlin: Springer.

Fischer, C. S. (2009). The 2004 GSS finding of Shrunken social networks: An artifact? *American Sociological Review, 74*, 657–669.

Flora, J. L., Green, G. P., Gale, E. A., Schmidt, F. E., & Flora, C. B. (1992). Self employment: A viable rural development option?". *Policy Studies Journal, 20*, 276–288.

Flora, J. L., Sharp, J., Flora, C., & Newlon, B. (1997). Entrepreneurial social infrastructure and locally initiated economic development in the nonmetropolitan United States. *Sociological Quarterly, 38*, 623–645.

Green, G. P. (2003). What role can community play in local economic development? In D. L. Brown & L. E. Swanson (Eds.), *Challenges for rural America in the twenty-first century* (pp. 343–353). University Park: Pennsylvania State University Press.

Green, D. P., & Gerber, A. S. (2008). *Get out the vote: How to increase voter turnout*. Washington, DC: Brookings Institution Press.

Irwin, M., Blanchard, T., Tolbert, C., Nucci, A., & Lyson, T. (2004). Why people stay: The impact of community context on nonmigration in the USA. *Population-E, 59*, 567–592.

Isserman, A., Feser, E., & Warren, D. E. (2009). Why some rural places prosper and others do not. *International Regional Science Review, 32*, 300–342.

Jobes, P. C. (1999). Residential stability and crime in small rural agricultural and recreational towns. *Sociological Perspectives, 42*, 499–524.

Johnson, K., & Rathge, R. W. (2006). Agricultural dependence and changing population in the great plains. In W. Kandel & D. Brown (Eds.), Population change and rural society (pp. 197–217). Dordrecht: Springer.

Kasarda, J. D., & Janowitz, M. (1974). Community attachment in mass society. *American Sociological Review, 39*, 328–339.

Lobao, L. M., Hooks, G., & Tickamyer, A. (2007). *The sociology of spatial inequality*. Albany: State University of New York Press.

McGranahan, D. (1999). *Natural amenities drive rural population change*. Agricultural Economic Report No. AER781, Economic Research Service, U.S. Department of Agriculture.

McGranahan, D. (2008). Landscape influence on recent rural migration in the U.S. *Landscape and Urban Planning, 85*, 228–240.

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in America: Changes in core discussion networks over two decades. *American Sociological Review, 71*, 353–375.

Mencken, F. C., Bader, C., & Polson, C. (2006). Integrating civil society and economic growth in Appalachia. *Growth and Change, 37*, 107–127.

Musick, M. A., Wilson, J., & Bynum, W. B. (2000). Race and formal volunteering: The differential effects of class and religion. *Social Forces, 78*, 1539–1570.

Perlmann, J. (2005). *Italians then, Mexicans now*. New York: Russell Sage.

Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Rathge, R. R., & Paula Highman, P. (1998). Population change in the great plains: A history of prolonged decline. *Rural Development Perspectives, 13*, 19–26.

Sampson, R. J. (1988). Local friendship ties and community attachment in mass society: A multilevel systemic model. *American Sociological Review, 53*, 766–779.

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social disorganization theory. *American Journal of Sociology, 94*, 774–802.

Schwarzweller, H. K., & Lean, S. (1993). Ontonagon: A remote corner of Michigan's Upper Peninsula. In: T. A. Lyson & W. W. Falk (Eds.), *Forgotten places* (pp. 168–194). Lawrence: University Press of Kansas.

Sondheim, R. M., & Green, D. P. (2010). Using experiments to estimate the effects of education on voter turnout. *American Journal of Political Science, 54*, 174–189.

Stoll, M. A. (2001). Race, neighborhood poverty, and participation in voluntary associations. *Sociological Forum, 16*, 529–557.

Terriquez, V. (2012). Civic inequalities? Immigrant incorporation and Latina Mothers' participation in their children's schools. *Sociological Perspectives, 55*, 663–682.

Tolbert, C. M. (2005). Minding our own business: Locally oriented businesses and the future of southern civic community. *Social Forces, 83*, 1309–1328.

Tolbert, C., Thomas Lyson, M., & Irwin, M. (1998). Local capitalism, civic engagement, and socioeconomic well-being. *Social Forces, 77*(2), 401–427.

Tolbert, C. M., Irwin, M. D., Lyson, T. A., & Nucci, A. R. (2002). Civic community in small town America: How civic welfare is influenced by local capitalism and civic engagement*. *Rural Sociology, 67*, 90–113.

Tomaskovic-Devey, D., & Roscigno, V. J. (1997). Uneven development and local inequality in the U.S. South: The role of outside investment, landed elites, and racial dynamics. *Sociological Forum, 12*, 565–597.

Wilson, J. (2000). Volunteering. *Annual Review of Sociology, 26*, 215–240.

Wimberley, R. C., & Morris, L. V. (2002). The regionalization of poverty: Assistance for the black belt south? *Southern Rural Sociology, 18*, 294–306.

Chapter 13
Revisiting the Rural Paradox in US Counties with Spatial Durbin Modeling

Tse-Chuan Yang, Aggie J. Noah, and Carla Shoff

13.1 Introduction

The standardized mortality rate is an important indicator of population health, and disparities in mortality have persisted along various dimensions including race/ethnicity, class, gender, and geographic space. One area in mortality research that has received considerable scholarly attention in the past few decades is rural-urban differentials in mortality. The so-called "rural paradox" refers to the phenomenon in which standardized mortality rates are unexpectedly lower in rural counties than in their urban counterparts, despite worse socioeconomic profiles and less health infrastructure in rural counties (Yang et al. 2011). Although rural counties have higher crude death rates, standardization for demographic composition (e.g., age and sex) reverses this rural disadvantage; and such advantages in standardized mortality rates are consistent regardless of the operationalization of the concept of "rural" (Yang et al. 2011; McLaughlin et al. 2001; McLaughlin et al. 2007). Rural sociologists, in particular, have extensively documented the rural paradox since the mid-1980s (Miller et al. 1987; Clifford et al. 1986; Clifford and Brannon 1985), and they have also persistently worked to explain the paradoxical phenomenon (McLaughlin et al. 2001, 2007; Yang et al. 2011). Previous studies have

T.-C. Yang (✉)
Department of Sociology, Center for Social and Demographic Analysis, University at Albany, State University of New York, New York, NY, USA
e-mail: tyang3@albany.edu

A.J. Noah
Department of Sociology and Criminology, Pennsylvania State University, State College, PA, USA

C. Shoff
Centers for Medicare and Medicaid Services, Office of Enterprise Data and Analytics, Baltimore, MD, USA

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_13

examined several plausible explanatory variables such as racial composition, social capital, and income inequality to better understand the salubrious influence of rural residence on mortality (Yang et al. 2011, 2012; McLaughlin et al. 2007; McLaughlin and Stokes 2002). For instance, the better social capital and lower income inequality in rural areas than urban have partly accounted for the rural paradox (Yang et al. 2011; McLaughlin et al. 2007). However, despite these scholarly efforts to identify the factors and potential mechanisms underlying the rural paradox, the advantage of rural residence on mortality has not yet been fully explained.

This current study identifies three major theoretical and methodological shortcomings of previous rurality-mortality research. First, the simplistic conceptualization of “rurality” is problematic. The term “rural” means more than where individuals live; and its theoretical root, rurality, is a multidimensional concept (Willits et al. 1990; Bealer et al. 1965; Miller and Luloff 1981). Previous rural sociological literature discusses how the concept of rurality should at least encompass ecological, occupational, and sociocultural dimensions (Bealer et al. 1965); however, most researchers have overemphasized the ecological dimension by only measuring rural based on the population size of a county. This approach simplifies rurality into a simple classification scheme, and may disguise the effects of other aspects of rurality (e.g., economic integration into adjacent areas and natural resource dependency) on mortality. Furthermore, the geographic development of rurality is not mutually independent, because counties that are nearby may share similar characteristics (e.g., natural resources, cultures, religions, etc.) and hence, have similar dimensions of rurality. Previous mortality research has paid relatively little attention to these issues regarding the conceptualization and operationalization of rurality.

Second, previous studies have not proposed a theoretical framework that situates the rural paradox into the geographic health disparity literature, despite the fact that there is a call for this research (Sparks et al. 2013). As mentioned earlier, the standardized mortality rate is an important indicator of population health, and the reduction in the mortality disparity is an important public health goal. The reduction in mortality results from a combination of factors, including improved standards of living, greater access to health care, and a clear understanding of risky factors (Hoyert 2012); and yet, rural-urban differentials in mortality do not fit into this conventional understanding of the relationship between protective/risk factors and mortality. A thorough theoretical framework will help elucidate the factors and potential mechanisms behind the rural paradox, and subsequently design public policy to systematically reduce the mortality disparity between urban and rural areas.

As the conceptualization and operationalization of rurality should incorporate the interdependence of rurality, a theoretical framework investigating the rural paradox should also examine the effects of the residential context beyond the immediate residential area (Dietz 2002; Mujahid and Diez Roux 2010). Most, if not all, previous mortality studies have used a micro-demography perspective (Voss 2007), and focused on the mortality rate of a given area *only* with the characteristics *within* this area. However, mortality in the US is spatially dependent (James et al. 2004; Yang et al. 2011), and neglecting to move beyond the typical theoretical framework that only focuses on the "immediate area" can undermine our understanding of the rural paradox.

Third, extending from the previous point, researchers have begun to realize that ecological mortality data are subject to spatial dependence, which requires the use of spatial regression methods to obtain unbiased coefficient estimates (Sparks et al. 2013; Sparks and Sparks 2010). Not accounting for spatial dependence in ecological mortality data can bias the results, lead to incorrect estimates of the associations between the independent and dependent variables, and ultimately, lead to improper conclusions (Haining 2003; Yang et al. 2011). The common practice is to model spatial dependence in either the dependent variable (spatial lag model) or the error structure (spatial error model) (Anselin 1988), and then use the features of a county to understand why rurality is negatively related to the mortality within a county. However, we argue that this approach limits the exploration of explanations of the rural paradox as researchers presume that the mortality of a county is *only* related to the features within that county. This presumption overlooks the fact that social processes are spatially embedded and the formation of social relationships occurs across the county boundary (Cho et al. 2012).

The current study aims to address these definitional, theoretical, and methodological shortcomings of previous mortality research by developing a theoretical framework for understanding the rural paradox and by examining it with appropriate spatial analysis techniques that account for spatial dependence and the interplays across county boundaries. The results of this study will not only contribute to the geographic health (mortality) disparity literature, but also to demographic research in general, as we will provide an example of how spatial dependence could inform both micro- and macro-demographers (Voss 2007) about the determinants of demographic outcomes.

13.2 Spatializing the Hypotheses for Exploring the Rural Paradox

Although there is no existing theory for explaining the rural paradox, two hypotheses from the health disparities literature are helpful for better situating the rural paradox in population health research: the drift hypothesis and the breeder hypothesis (Verheij 1996). The drift hypothesis describes the geographic variations in health outcomes as a consequence of the fact that ill or susceptible individuals move to particular places and remain there; the breeder hypothesis, meanwhile, indicates that the geographic variations are the result of different exposures to environmental factors or different clusters of behaviors across space (Verheij 1996). Therefore, on the one hand, according to the drift hypothesis, it is possible that ill or susceptible residents move from counties with high rurality to those with low rurality and then die there, leading to the beneficial effect of rurality on mortality. For example, elderly individuals may relocate to more urban areas in order to utilize better health infrastructure (e.g., hospital or nursing homes) and transportation as they begin to encounter moderate chronic disability (Wilmoth 2010). On the other hand, extending the breeder hypothesis, it is possible that the residents in counties with high rurality have lifestyles, health behaviors, and environments that ameliorate their general health, and in turn, contribute to the rural paradox. It is important to note that these hypotheses are inherently spatial, as the drift hypothesis focuses on the interplays across areal boundaries and the breeder hypothesis emphasizes the importance of similarities within and between places. Thus, incorporating spatial perspectives into these competing theories is imperative.

Drawing spatial perspectives from other disciplines, we can further strengthen these two theoretical arguments. Specifically, the spatial spillover perspective (Cho et al. 2012) dictates that if a county could not afford a certain number of patients, then residents may relocate to seek health care elsewhere (i.e., spillover to neighbors). This spatial spillover perspective echoes the drift hypothesis. That is, the movement of ill individuals across space changes the composition within the area, and generates geographic variation in mortality (i.e., spillover of ill or susceptible individuals to particular places). This perspective can also be applied to other explanatory covariates. For instance, should the labor market be great in one particular county, residents in nearby counties may benefit from the economic opportunities that directly relate to human health (Link and Phelan 1995).

Alternatively, previous studies have found that people who pursue a healthy life and live nearby may have limited access to various resources, and they may compete with one another to secure the limited resources (Ginther et al. 2000; Firebaugh and Schroeder 2009). Applying this argument to the breeder hypothesis, people from the counties with high rurality (e.g., low population) may have relatively low competition for resources (e.g., natural amenities) within counties, however, the residents in nearby counties with low rurality may increase the level of competition for resources, then leading to the uneven geographic mortality distribution. It is also likely that residents in rural areas are less likely to be surrounded by areas with a significantly better socioeconomic status (i.e., more homogenous socioeconomic status across rural counties), and are less likely to experience relative deprivation; which, in turn, explains the rural paradox.

The drift and breeder hypotheses help us to better understand the rural paradox, and highlight the importance of the spatial relationships of a given county with its neighbors. These hypotheses will help elucidate the potential mechanisms for the rural paradox. In order to properly examine these two hypotheses and explicitly consider spatial neighbors in our analyses, we will employ spatial Durbin modeling (Anselin 1988) as the key tool in future analysis and compare spatial Durbin modeling with other conventional spatial regression models. No previous studies have used spatial regression modeling to test these two hypotheses; our analysis will not only shed new light on whether and how the drift and breeder hypotheses may help explain the rural paradox, but also demonstrate how spatial Durbin modeling improves the model fit compared to other spatial methods. In addition, we partition the spatial Durbin results to spatially profile the importance of neighbors. This study is among the first ecological mortality studies that attempt to use this spatial modeling perspective, e.g., Yang et al. (2015).

13.3 Methodology

13.3.1 Data and Measures

Following the discussion above, we calculated the 1998–2002 (5-year) average mortality rates standardized with the 2000 US age-sex population structure as the dependent variable (deaths/per 1000 population). The Compressed Mortality Files (CMF) are the major data sources maintained by the National Center for Health Statistics (NCHS 2011) and the purpose of standardization is to allow for comparisons across counties (Preston et al. 2001). While it is viable to further consider racial/ethnic structure in the standardization process, the CMF only categorizes race/ethnicity into white, black, and other races. This trichotomization ignores the growing Hispanic population and thus, we opt not to standardize by race/ethnicity. Instead, we include three racial/ethnic composition variables (see below) in our analysis in order to control for the impact of racial/ethnic structure on mortality.

Our independent variables could be classified into seven groups as follows:

Rurality. As discussed previously, rurality is a complex concept and there is no consensus on the measurement of rurality. While the rural-urban continuum codes and the urban influence codes are commonly used in mortality research (Yang et al. 2011; Morton 2004), these measurements focus mainly on population size, which is traditionally referred to as the ecological dimension of rurality (Bealer et al. 1965). In order to better reflect the complexity of rurality, we first derived six indicators from the 2000 Census and applied principal component analysis (PCA) to them. The PCA results indicated that rurality can be captured with three factors, which echoes the literature suggesting that rurality is multidimensional (Willits et al. 1990; Bealer et al. 1965).[1]

Specifically, the first factor of rurality refers to the *ecological dimension* that is related to the total number of population in a county. Three variables were highly loaded on this factor: population density (total population divided by total land area, factor loading =0.93), road density (the length of major roads divided by total land area, 0.80), and percentage of workers commuting with public transportation (0.95). Higher factor scores in ecological dimension indicate lower rurality. The second factor is *economic integration*, which includes two indicators, percentage of workers traveling over one hour to work (0.87) and

[1] Brown and Schafft (2011) discussed how rural residents are affected by the change in rural economies, institutions, and environment. The rurality measures of this study largely align with their perspective, but their organizational aspect of rurality is not considered in the analysis due to data limitations.

percentage of workers employed outside their county of residence (0.82). As this factor shows high factor scores in counties near metropolitan areas, a high level of economic integration indicates low rurality. The third factor is *natural resources dependency*, with one highly loaded indicator: percent of the population employed in farming, forestry, and fishing (0.93). Counties that are more dependent on natural resources are defined as having a higher level of rurality.

Racial/ethnic compositions. Drawing from the 2000 Census, we created three variables to capture the racial/ethnic composition of a county: the *percentage of non-Hispanic black*, the *percentage of Hispanics*, and the *percentage of non-Hispanic other races* (including multiple races). The reason why the percentage of non-Hispanic white was not included is to avoid perfect multicollinearity in regression analysis.

Migration. The drift hypothesis suggests that the rural paradox may be the consequence of internal migration. To test this hypothesis, we obtained the County-to-County Migration Data from the US Census Bureau and created two variables based on the reported residence 5 years prior to 2000: percent of elderly in-migration (aged 55 and over) and percent of young in-migration (aged 20 to 29). It should be noted that these two internal migration flows may impose opposite impacts on mortality. Specifically, the elderly in-migration may be positively associated with mortality, but the young in-migration may be negatively related to mortality. To better assess the overall impact of in-migration flows, we generated a single *in-migration flow* variable by subtracting the percentage of young in-migration from the percentage of elderly in-migration. Should the drift hypothesis stand, this in-migration flow variable would be positively related to mortality, net of other covariates.

Socioeconomic status (SES). As Link and Phelan (1995) have suggested, social conditions should be regarded as the fundamental determinants of health. In order to capture the socioeconomic status of a county, we followed the approach developed by Sampson et al. (1997) by extracting eight variables from the 2000 Census and conducting PCA to generate two variables based on the factor scores, namely *social affluence* and *concentrated disadvantage*. The former included the log of per capita income (factor loading = 0.88), percentage of the population age 25 or older with at least a bachelor's degree (0.93), percentage of the population employed in professional, administrative, and managerial positions (0.78), and the percentage of families with an income over $75,000 (0.92). The latter was loaded by poverty (0.89), percentage of persons receiving public assistance (.85), unemployment (.87), and the percentage of female-headed households with children (.78). These SES measures have been found to be strongly related to mortality in recent studies (Yang et al. 2011).

Income inequality. In addition to the absolute social conditions captured by the SES measures discussed above, the relative social conditions were also considered in this study. Several recent studies found that *income inequality*, a relative measure of wealth, is positively related to mortality (Kawachi et al. 1999; Kawachi and Kennedy 1999; Yang et al. 2012). We used the Gini coefficient to measure income inequality, which ranges between 0 (completely equal distribution) and

1 (completely unequal). The choice of measurements of income inequality has been found to be unrelated to the relationship between mortality and income inequality (Kawachi and Kennedy 1997). Using the household income data in 2000 Census, we calculate the Gini coefficient for all contiguous counties.

Social capital. To thoroughly examine the breeder hypothesis, the concept of social capital is also considered in the analysis (Putnam 2001). Using county level mortality data, a recent study suggested that social capital helps to explain the geographical mortality differential in the US in which higher social capital is associated with lower mortality rate (Yang et al. 2011). Rupasingha and colleagues (2006) developed a county-level *social capital index* based on the following four indicators: total number of civic organizations per 1000 population, total number of tax-exempt non-profit organizations per 1000 population, the 2000 Census response rate, and the presidential voting rate in 2000. These data are publicly available (Rupasingha and Goetz 2008). While there are many definitions of social capital in the existing literature (Song et al. 2010), a recent study suggested that this social capital index is the best available measurement of social capital at the county level (Shoff and Yang 2013).

Environmental hazards. The variables above are all related to social conditions and factors for mortality; however, ecological mortality research has overlooked the impact of environmental hazards. To fill this gap, three variables were created with the data from the US Environmental Protection Agency (EPA), including the Toxics Release Inventory (TRI) and the Air Quality System (AQS). TRI provides information on chemical releases and waste management related to manufacturing facilities. For each county, *toxics density* was obtained by dividing the total chemical releases (in pounds) by the total land area. In addition, the total amount of carcinogenic chemicals in TRI were further converted into pounds of benzene-equivalents[2] and divided by the total land area, which creates the second variable, *density of TRI-related carcinogens*. AQS includes the *air quality index (AQI)* that measures five common pollutants: particulate matter, sulfur dioxide, carbon monoxide, nitrogen dioxide, and ozone. High AQI values represent high levels of these five pollutants. In order to minimize the variation among the three variables, we standardized them into variables with a mean of 0 and a variance of 1.

[2] Benzene-equivalents indicate the amounts of benzene that would have to be released into the air to pose the same level of health risk as the release of other chemicals. It is a useful measure to compare different carcinogenic toxic releases and their risks to benzene. The list of carcinogen chemicals could be found in the website below: http://www2.epa.gov/sites/production/files/documents/OSHA_carcinogen_table_2011.pdf

13.3.2 Spatial Durbin Modeling

Extending from Voss (2007), we argue that spatial analysis should be treated as the "conventional" analytic approach when handling ecological demographic data. Spatial econometrics modeling has been widely used in the spatial social sciences (LeSage and Pace 2009) and spatial lag and spatial error models are commonly used (Elhorst 2010). While these models allow researchers to control for spatial dependence, they may not answer how the spatial structure underlying the data matters. Elhorst (2010) further suggested that it is time to shift from these models to the spatial Durbin model (Anselin 1988), because the latter has been proven to outperform the others. Specifically, the spatial Durbin model is "the only means of producing unbiased coefficient estimates," regardless of the true spatial process underlying the observed data (Elhorst 2010). That is, without any prior knowledge of how the data are related spatially (e.g., mortality and rurality over counties), researchers could still obtain unbiased estimates and make correct inferences. This advantage helps to test our two hypotheses as they have never been examined in the rural paradox literature, and using the spatial Durbin will avoid severe consequences of model misspecifications. This study will be among the first to use a spatial Durbin model to answer the questions of whether and how the drift and breeder hypotheses explain the rural paradox.

Technically, a spatial Durbin model is comprised of three components: a spatial lagged dependent variable, a set of explanatory variables of a spatial unit, and a set of spatial lagged explanatory variables (LeSage and Pace 2009), which can be expressed as:

$$
\begin{aligned}
&y = \rho W y + \alpha l_n + X\beta + WX\theta + \varepsilon \\
&\varepsilon \sim N(0, \sigma^2 I_n),
\end{aligned}
$$

where y denotes an n × 1 vector of the dependent variable (i.e., mortality), W is the n × n spatial weight matrix, Wy represents the spatial lagged dependent variable, ρ denotes an 1 × n vector of the effects of Wy, and l_n indicates an n × 1 vector of ones associated with the intercept parameter α. X represents an n × k matrix of k explanatory variables, which are related to the parameters β; WX reflects the spatial lagged explanatory variables and θ denotes a k × 1 vector of the effects of WX. The error tem, ε, follows a normal distribution with a mean of 0 and a variance $\sigma^2 I_n$, where I_n is an n × n identify matrix.

The model above explicitly takes into account the exogenous interactions ($WX\theta$) between the mortality of a particular county and the features of its neighbors, as well as the endogenous relationships between the mortality and explanatory covariates within a county ($X\beta$). Though the error term above could be further divided into spatially structured and random errors, doing so would generate biased estimates of the relationships between the dependent variable and the independent variables (Manski 1993; Elhorst 2010). Therefore, the formula above could be regarded as the most appropriate model for this study.

13.3.3 Analytic Strategy

To understand whether and how the drift and breeder hypotheses help to explain the rural paradox, our analytic strategy consists of three stages. First, we implement five regression models. The baseline model only considers rurality and racial compositions, as we need to understand if the rural paradox exists in our data. To examine the drift hypothesis, we add internal migration into the baseline model as our second model. Similarly, the third model adds SES, income inequality, social capital, and environmental hazards into the baseline model. Internal migration is added to the third model to create the fourth model where we would like to find evidence for both the drift and breeder hypotheses. The last model further considers the exogenous relationships between the mortality rate of a county and the features in its neighboring counties (spatial Durbin model).

The second stage of the analysis focuses on decomposing the effects of the explanatory variables in the last model into direct and indirect impacts. As LeSage and Pace (2009) suggested, another major advantage of applying a spatial Durbin model to empirical studies is to allow researchers to thoroughly understand the spatial dynamics between the dependent and independent variables across space. More specifically, the effects found in a spatial Durbin model could be dichotomized into direct and indirect impacts. The former is to capture how the change in one independent variable is associated with the change in the dependent variable *within* a spatial unit, whereas the latter quantifies the overall impact of how the change in one independent variable of a spatial unit affects the dependent variables in *other* spatial units.

Our third stage is to further partition the direct and indirect impacts by neighboring orders, such as first-order (immediate) neighbors and second-order neighbors (neighbors of the immediate neighbors). Due to the space constraint, we opt not to discuss the technical details of this stage, which could be found in LeSage and Pace (2009). The partitioning results would spatially profile the importance of neighbors, and help us to better understand if the spatial structure really contributes to the rural paradox. Note that the partitioning process has not been commonly used until recently (Autant-Bernard and LeSage 2011; Jensen et al. 2012).

13.4 Results

Table 13.1 presents the results of the five regression models described above. The key findings are summarized as follows. First, the baseline model (Model 1 in Table 13.1) identified the complexity of the rural paradox (Yang et al. 2011) and suggested that the relationship between rurality and mortality may vary by the rurality dimensions. Specifically, the ecological dimension of rurality was negatively related to mortality, indicating that the age-sex standardized mortality rates increase with population. This does not confirm the rural paradox. However, the

Table 13.1 Spatial lag and Durbin regression results

	Model 1		Model 2		Model 3		Model 4		Model 5			
	Estimate		Estimate		Estimate		Estimate		Estimate		Lag Estimate	
Intercept	3.710	***	3.712	***	5.461	***	5.377	***	4.995	***		
Rurality												
Ecological dimension	−0.046	***	−0.045	***	0.066	***	0.063	***	0.014			
Economic integration	0.101	***	0.100	***	0.060	***	0.068	***	0.020			
Natural resources dependency	−0.062	***	−0.063	***	−0.197	***	−0.196	***	−0.128	***		
Racial composition												
% Black	0.025	***	0.025	***	0.010	***	0.010	***	0.007	***		
% Hispanic	0.000		0.000		−0.012	***	−0.012	***	−0.015	***		
% Others	0.040	***	0.040	***	0.015	***	0.015	***	0.015	***		
SES												
Social affluence					−0.335	***	−0.340	***	−0.387	***	0.134	**
Concentrated disadvantage					0.345	***	0.338	***	0.471	***	−0.361	***
Migration												
Internal migration			0.043				−0.264		−0.163		−0.766	**
Inequality												
Income inequality					0.363		0.570		0.535		0.445	
Social capital												
Social capital index					−0.087	***	−0.088	***	0.016		−0.309	***

Environmental hazards											
Standardized air quality index					0.012		0.014		0.012		−0.033
Toxic density					0.041	*	0.041	*	0.029		0.084
Standardized TRI-related carcinogens					0.024		0.024		0.019		−0.033
Spatial effect											
Rho (Spatial Lag)	0.544	***	0.544	***	0.362	***	0.360	***	0.383	***	
Model diagnostics											
AIC	8956.8		8958.7		8152.3		8151.6		7963.6		
Likelihood ratio test	789.3	***	789.2	***	322.1	***	317.5	***	267.6	***	

Note: $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$

other two dimensions of rurality, economic integration and natural resources dependency, demonstrated that counties that are more economically isolated or dependent on farming, fishing and forestry tend to have lower mortality rates than their counterparts. These associations of rurality with mortality are net of the impacts of racial compositions in a county. Second, Model 2 provides evidence for researchers to understand whether internal migration explains the rural paradox, but we did not find a statistically significant relationship between internal migration and mortality. Although the direction of the impact of internal migration on mortality follows the theoretical expectation, including this variable did not alter how rurality is associated with mortality. Even after controlling for other social conditions (see Model 4), internal migration remains a non-significant factor for age-sex standardized mortality rates.

Third, echoing the argument that social conditions are the fundamental determinants of health (Link and Phelan 1995), our Models 3 and 4 suggested that SES and social capital are not only associated with mortality, but also affect the rural paradox. Counties with higher scores of social affluence (e.g., residents with high educational attainment and income) are more likely to have lower mortality than those with lower scores. This relationship was also verified by concentrated disadvantage (e.g., poverty and unemployment rate). In addition, the social capital index is negatively associated with mortality, suggesting that strong social connections among residents benefit population health in terms of mortality. This protective impact of social capital on health has been found in recent studies (Kawachi et al. 2008; Song et al. 2010). Beyond social conditions, we also found that toxic density in a county was positively related to mortality, confirming that environmental hazards matter in determining the geographic mortality differentials in the contiguous US.

Including social conditions and environmental hazards in the analysis changes the relationships between rurality and mortality. In contrast to Model 1, the magnitude of the effect of economic integration on mortality decreased by over 30 % in Models 3 and 4. In addition, the direction of the association of the ecological dimension with mortality changed and suggested the existence of the rural paradox, with larger populations related to high mortality. As for natural resource dependency, its association with mortality became stronger from Model 1 to Model 3. These findings indicate that the rural paradox could be explained partly by social conditions and environmental hazards and that some of the paradox may be more paradoxical.

Fourth, the mixed results from Model 1 to Model 4 may have resulted from the fact that the features of neighboring counties are not considered. Using the spatial Durbin approach, we found that the rural paradox could be explained by the spatial structure underlying our data. To be specific, after taking the characteristics of neighboring counties into account (see Model 5), the ecological dimension and economic integration were no longer significantly related to mortality, and the magnitude of the effect of natural resources dependency dropped by roughly 35 % ((0.196–0.128)/0.196), albeit remaining significant. That said, the spatial Durbin model suggested that the so-called rural paradox stands only when rurality is measured by natural resources dependency and that the ecological dimension and economic integration may not contribute to the mortality disparities across space. Furthermore, comparing Model 5 with Model 4, the effects of social capital index and toxic density on mortality became non-significant within a county.

As for whether the spatial Durbin model overfits our data, we compared the Akaike Information Criterion (AIC) across the five models. We found that Model 5 has the lowest AIC and the differences between Model 5 and the others are all greater than 10, which indicates that the spatial Durbin model is the most preferred among the five models (Burnham and Anderson 2002). As AIC is an indicator that takes the total number of parameters into account, overfitting should not be a concern in our spatial Durbin model.

Following our analytic strategy, we decomposed the estimates in Model 5 into direct and indirect impacts. Similar to Autant-Bernard and LeSage (2011), we only focused on how the changes in the independent variables in a county affect its own (direct) mortality and the mortality in other counties (indirect). The decomposition results were summarized in Table 13.2. The total impacts in Table 13.2 are not necessarily equal to the estimates in Model 5, because the direct impacts refer to the diagonal elements of the n x n matrix (please see the methodology section) and the indirect impacts are represented by the off-diagonal elements. We refer interested readers to LeSage and Pace (2009) for technical details/explanations.

Several findings are notable in Table 13.2. We found that a one unit increase in the natural resources dependency score would lead to a decrease of 0.13 deaths per 1000 population within a county, and this same change in rurality would also lead to an overall decrease of 0.08 deaths per 1000 population across the contiguous US. The direct impact is roughly 1.7 times larger than the indirect impact, while both are statistically significant. Moreover, the findings of racial compositions corresponded to the literature (McLaughlin and Stokes 2002). The percentages of the non-Hispanic black population and non-Hispanic other races are positively associated with mortality and their direct impacts are about twice as strong as the indirect impacts. Generally, a one percent increase in these variables is associated with an increase of about 0.02 (0.016 + 0.008) deaths per 1000 population within a county, and roughly 0.01 (0.004 + 0.009) deaths per 1000 population in other counties. The negative relationship of the percent of Hispanics with mortality seems to support the so-called “Hispanic paradox” in the mortality literature (Abraido-Lanza et al. 1999), which states that Hispanics have lower mortality compared to other racial/ethnic groups despite their disadvantaged socioeconomic status, and the magnitudes of the direct and indirect impacts are comparable with those of non-Hispanic other races.

As for the SES variables, their impacts on mortality are quite profound. Within a particular county, a one-unit increase in the social affluence score would reduce mortality by 0.4 deaths per 1000 population; moreover, the change in social affluence would spill over to other counties and yield an overall decrease of 0.2 deaths per 1000 population across the US. Concentrated disadvantage imposed a stronger impact on mortality than social affluence, with a one-unit increase

Table 13.2 Decomposition estimates of the direct and indirect effects of selected conditions on mortality

	Total	Direct	Indirect
Rurality			
Ecological dimension	0.022	0.014	0.008
Economic integration	0.033	0.021	0.012
Natural resources dependency	**−0.208**	**−0.132**	**−0.076**
Racial compositions			
% Black	**0.012**	**0.008**	**0.004**
% Hispanic	**−0.024**	**−0.015**	**−0.009**
% Others	**0.024**	**0.016**	**0.009**
SES			
Social affluence	**−0.626**	**−0.399**	**−0.228**
Concentrated disadvantage	**0.762**	**0.485**	**0.277**
Migration			
Internal migration	−0.263	−0.168	−0.096
Inequality			
Income inequality	0.866	0.551	0.315
Social capital			
Social capital index	0.025	0.016	0.009
Environmental hazards			
Standardized air quality index	0.020	0.012	0.007
Toxic density	0.047	0.030	0.017
Standardized TRI-related carcinogens	0.031	0.020	0.011

Note: Bold face type indicates that the variable is associated with the dependent variable at the 95 % level

Note: $^*p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$

associated with increases of 0.5 and 0.3 deaths per 1000 population in mortality within and across counties, respectively.

Using the infinite expansion technique (LeSage and Pace 2009), we further profiled the direct and indirect impacts spatially (by neighboring orders) and the results are presented in Table 13.3. This partitioning technique assumes that the indirect effects are null for zero-order (within counties) neighbors and that the direct effects are null for first-order (immediate) neighbors (Autant-Bernard and LeSage 2011). According to the partitioning results, we found evidence for spatial feedback effects. That is, for direct impacts, the significant impacts beyond the second-order neighboring (i.e., W_2, W_3, and W_4) could be understood as the fact that a county is a second-order (or higher) neighbor to itself (LeSage and Pace 2009), which contributes to the diagonal elements (direct effect). For instance, this phenomenon was observed for natural resources dependency, social affluence, and concentrated disadvantage. While we can list the results for the neighboring order higher than 4, they are relatively trivial and we opted not to include them in the table (available upon request).

Table 13.3 Spatial partitioning results of direct and indirect effects of selected conditions on mortality

	Direct					Indirect				
	W_0	W_1	W_2	W_3	W_4	W_0	W_1	W_2	W_3	W_4
Rurality										
Ecological dimension	0.014	0.000	0.000	0.000	0.000	0.000	0.005	0.002	0.001	0.000
Economic integration	0.020	0.000	0.001	0.000	0.000	0.000	0.008	0.002	0.001	0.000
Natural resources dependency	**−0.128**	0.000	**−0.003**	**0.000**	**0.000**	0.000	**−0.049**	**−0.016**	**−0.007**	**−0.003**
Racial compositions										
% Black	**0.007**	0.000	**0.000**	**0.000**	**0.000**	0.000	**0.003**	**0.001**	**0.000**	**0.000**
% Hispanic	**−0.015**	0.000	**0.000**	**0.000**	**0.000**	0.000	**−0.006**	**−0.002**	**−0.001**	**0.000**
% Others	**0.015**	0.000	**0.000**	**0.000**	**0.000**	0.000	**0.006**	**0.002**	**0.001**	**0.000**
SES										
Social affluence	**−0.387**	0.000	**−0.010**	**−0.001**	**−0.001**	0.000	**−0.148**	**−0.047**	**−0.020**	**−0.008**
Concentrated disadvantage	**0.471**	0.000	**0.012**	**0.001**	**0.001**	0.000	**0.180**	**0.057**	**0.025**	**0.009**
Migration										
Internal migration	−0.163	0.000	−0.004	−0.001	0.000	0.000	−0.062	−0.020	−0.009	−0.003
Inequality										
Income inequality	0.535	0.000	0.013	0.002	0.001	0.000	0.205	0.065	0.028	0.011
Social capital										
Social capital index	0.016	0.000	0.000	0.000	0.000	0.000	0.006	0.002	0.001	0.000
Environmental hazards										
Standardized air quality index	0.012	0.000	0.000	0.000	0.000	0.000	0.005	0.001	0.001	0.000
Toxic density	0.029	0.000	0.001	0.000	0.000	0.000	0.011	0.004	0.002	0.001
Standardized TRI-related carcinogens	0.019	0.000	0.000	0.000	0.000	0.000	0.007	0.002	0.001	0.000

Note: Bold face type indicates that the variable is associated with the dependent variable at the 95 % level

As for the indirect impacts, one common pattern is that the magnitudes of the effects on mortality decayed rapidly between the first- and second-order neighbors. For instance, the indirect impact of natural resources dependency on mortality dropped by almost 70 % from first- to second-order neighbors, then decreased by another 55 % to the third-order, and finally diminished to −0.003, which is only 6 % of the impact on the first-order neighbors. Coupled with the SES variables, the partitioning results suggested that the indirect impacts should be the most profound on the immediate neighbors, and that beyond the third-order neighbors, the indirect impacts become quite small. By contrast, the impacts of racial compositions seemed to be less relevant to geographic proximity as their spatial feedback effects (direct) were largely concentrated on zero-order and their partitioned indirect impacts were mostly bounded by the second-order neighbors. Our partitioning of the direct and indirect impacts by neighboring orders better depicts what the spatial dynamic processes look like among counties and should provide a clearer answer to the question of how the spatial structure matters in ecological mortality research.

13.5 Discussion and Conclusions

The rural paradox refers to the fact that despite the poor socioeconomic profiles of rural residents, the standardized mortality rates are lower in rural than urban areas (Yang et al. 2011). While this phenomenon has been previously documented (Miller et al. 1987; Clifford et al. 1986), little is known about how the spatial structure underlying the ecological data matters and whether the drift and breeder hypotheses explain the rural paradox. Moreover, previous county-level mortality studies did not pay much attention to the complexity of the concept of rurality. This study contributes to the literature by filling these gaps.

Measuring rurality with the ecological dimension, economic integration, and natural resources dependency, we found that the rural paradox stands for all three dimensions of rurality when the characteristics of neighboring counties were not considered. The inclusion of internal migration in the analysis did not explain the impacts of rurality dimensions on mortality, which leads us to conclude that the drift hypothesis did not contribute to the rural paradox. To further verify this conclusion, we included the percent of elderly in-migration and the percent of young in-migration into the analysis, but the results did not change. We note that our dependent variable has been age-sex standardized and this may be the reason why age-specific internal migration did not matter. Though we could use crude death rates as our dependent variable, doing so would not justify the comparisons of mortality rates across different areas (Preston et al. 2001).

While we found no sufficient support for the drift hypothesis, adding the social conditions and environmental hazards measures seemed to provide evidence for the breeder hypothesis to some extent. These covariates explained almost 40 % of the association between economic integration and mortality, suggesting that the characteristics of where people live contribute to the rural paradox in terms of the economic perspective. Explicitly, one reason for the geographic mortality differentials in the US is the uneven distribution of social and environmental factors across space. However, the associations of other dimensions of rurality with mortality could not be understood in the same fashion.

Given the strengths of the spatial Durbin modeling (Elhorst 2010), we took the characteristics of neighboring counties into account to understand if the spatial structure contributes to the rural paradox and if so, how it matters. The relationships of the ecological dimension and economic integration with mortality became statistically non-significant when the independent variables of neighbors were considered, though natural resources dependency remained a significant determinant. That being said, to some degree, the rural paradox may be a consequence of the spatial dynamic processes embedded in our data. The traditional analytic approach (i.e., ordinary least squares regression) and the commonly used spatial lag model fail to incorporate the exogenous relationships between the mortality rate of a county and the features of its neighbors. Our analytic results suggested that among the three dimensions of rurality, the rural paradox could only be observed in natural resources dependency. An earlier study concluded that the percent of population working in agriculture was negatively related to mortality at the county-level (McLaughlin et al. 2007). Our finding echoed this conclusion and, moreover, our evidence was stronger, as both the spatial structure and environmental hazards were included in the analysis.

Furthermore, the decomposition and partitioning results advance the rural paradox literature by challenging the existing perspective of handling the spatial structure. Explicitly, current spatial mortality research has largely overlooked the interactions between the dependent variable of a unit and the independent variables of its neighbors. To our knowledge, the spatial Durbin approach has not been applied to the US county-level data to investigate geographic mortality disparities. In general, our results suggested that the total impacts of the independent variables on mortality could be further divided into direct impacts (roughly 65 %) and indirect impacts (about 35 %). While the spatial feedback effects (direct) were relatively trivial, they were still observed due to the dynamics (interactions) among counties. More importantly, most of the spatial dynamic processes do not go beyond the second-order neighbors, while the indirect impacts of SES variables on mortality remained moderate at the third-order. The decomposition and partitioning results convey an imperative message: that arbitrarily defined county boundaries do not limit human and social behaviors among residents. At least, the overall assessment of the health of the population in a county, namely mortality, is not merely determined by the characteristics of the population, but also by the interactions between the local population and their neighbors. By considering the interactions across space, the rural paradox is partly dismantled.

While this study advances our understanding of geographic mortality disparities in the US, several limitations should be noted. First, the spatial dynamic processes are assumed to be static, because the data are cross-sectional and centered on 2000. It is possible that the interactions among counties may change over time as human behaviors, attitudes, and environmental factors are not time-invariant and could be captured and analyzed with other techniques should the data support them (Elhorst 2010). Second, while this study has utilized the spatial Durbin approach to model the exogenous relationship (Elhorst 2010), it is still subject to the modifiable area unit problem (Openshaw 1983; Fotheringham and Wong 1991). That said, should the death data be aggregated into different geographic boundaries, our findings and conclusions may be altered. Third, our measure of internal migration may not fully reflect the drift hypothesis (Verheij 1996), as this hypothesis argues that the migration of unhealthy people leads to the geographic health disparities. Our age-specific internal migration variables may not be able to precisely reflect the health status of the in-migrants. Similarly, though our measures of environmental hazards are maintained by federal agencies, they do not precisely capture individuals' exposure to toxics, air pollutants, and carcinogens. Fourth, due to the space constraint, this study focuses heavily on how the changes in the independent variables of a county affect the mortality of its own and of its neighbors. Relatively little attention has been paid to the lagged estimates. However, as the leading spatial econometricians suggest (Elhorst 2010; Autant-Bernard and LeSage 2011; LeSage and Pace 2009), the interpretations of the lagged independent variables may not be of primary interest and could further complicate the substantive findings. Finally, it should be noted that spatial Durbin modeling may be subject to multicollinearity issues, which would only affect the significance testing results and would not lead to unbiased estimates (LeSage and Pace 2009). Given the high significance of our analytic results, multicollinearity may not undermine this study.

Some policy implications and future research directions can be drawn from our findings. First, our finding that people living in a county with high natural resource dependency tend to be healthy implies that the work related to natural resources involves more physical activity than other industrial sectors, which would further promote population health (Bouchard et al. 2012). Extending this finding, encouraging people to engage in various physical activities, such as walking or gardening, should help reduce the mortality rate. Second, the crucial role of SES in mortality highlights the importance of education and access to health information. Thus, it becomes important for health education programs to reach out to high-risk populations in order to minimize the total number of preventable deaths. Third, this study demonstrated how changes in one county might lead to changes in other counties. Public health policy makers and researchers should not be confined to arbitrary administrative boundaries (Matthews 2011). For instance, the evaluation of health policies or interventions should not be limited to a certain area or population as residents nearby may benefit from these interventions or policies via interactions. Finally, in order to move beyond the suggestions provided by Voss (2007) and Elhorst (2010), future ecological social studies should recapture the associations of societies with geographic space by explicitly using the spatial structure, as well as other spatial concepts (e.g., proximity) to explain the variations of the dependent variable (Porter and Howell 2012).

Acknowledgements This study received support from the Geographic Information Analysis Core at Penn State's Population Research Institute, which receives core funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD; R24-HD41025). We also acknowledge the help from Family Demography Training (T-32HD007514) from NICHD.

References

Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the "salmon bias" and healthy migrant hypotheses. *American Journal of Public Health, 89*(10), 1543–1548.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Autant-Bernard, C., & LeSage, J. P. (2011). Quantifying knowledge spillovers using spatial econometric models. *Journal of Regional Science, 51*(3), 471–496.

Bealer, R. C., Willits, F. K., & Kuvlesky, W. P. (1965). The meaning of rurality in American society: Some implications of alternative definitions. *Rural Sociology, 30*(3), 255–266.

Bouchard, C., Blair, S. N., & Haskell, W. L. (2012). *Physical activity and health*. Champaign: Human Kinetics Publishers.

Brown, D. L., & Schafft, K. A. (2011). *Rural people and communities in the 21st century: Resilience and transformation*. Malden: Polity Press.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. New York: Springer.

Cho, S., Kim, J., Roberts, R., & Kim, S. (2012). Neighborhood spillover effects between rezoning and housing price. *The Annals of Regional Science, 48*(1), 1–19.

Clifford, W. B., & Brannon, Y. S. (1985). Rural-urban differentials in mortality. *Rural Sociology, 50*(2), 210–224.

Clifford, W. B., Miller, M. K., & Stokes, C. S. (1986). Rural-urban differences in mortality in the United States, 1970–1980. In *New dimensions in rural policy. Building upon Our heritage*. Washington, DC: GPO.

Dietz Roux, D. (2002). The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research, 31*(4), 539–575.

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis, 5*(1), 9–28.

Firebaugh, G., & Schroeder, M. B. (2009). Does your neighbor's income affect your happiness? *American Journal of Sociology, 115*(3), 805–831.

Fotheringham, A. S., & Wong, D. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A, 23*(7), 1025–1044.

Ginther, D., Haveman, R., & Wolfe, B. (2000). Neighborhood attributes as determinants of children's outcomes: How robust are the relationships? *Journal of Human Resources, 35*(4), 603–642.

Haining, R. (2003). *Spatial data analysis: Theory and practice*. Cambridge: Cambridge University Press.

Hoyert, D. L. (2012). 75 years of mortality in the United States, 1935–2010. *NCHS data brief* (Vol. 88). Hyattsville: National Center for Health Statistics.

James, W. L., Cossman, R. E., Cossman, J. S., Campbell, C., & Blanchard, T. (2004). A brief visual primer for the mapping of mortality trend data. *International Journal of Health Geographics, 3*(7), 1–17.

Jensen, C. D., Lacombe, D. J., & McIntyre, S. G. (2012). A Bayesian spatial econometric analysis of the 2010 UK General Election. *Papers in regional science*, DOI: 10.1111/j.1435-5957.2012.00415.x.

Kawachi, I., & Kennedy, B. P. (1997). The relationship of income inequality to mortality: Does the choice of indicator matter? *Social Science & Medicine, 45*(7), 1121–1127.

Kawachi, I., & Kennedy, B. P. (1999). Income inequality and health: Pathways and mechanisms. *Health Services Research, 34*(1 Pt 2), 215–227.

Kawachi, I., Kennedy, B. P., & Wilkinson, R. G. (1999). *Income inequality and health* (Vol. 506). New York: The New Press.

Kawachi, I., Subramanian, S. V., & Kim, D. (2008). Social capital and health: A decade of progress and beyond. In I. Kawachi, S. V. Subramanian, & D. Kim (Eds.), *Social capital and health* (pp. 1–26). New York: Springer.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics* (Vol. 196). Boca Raton: Chapman & Hall/CRC.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior, 35*, 80–94.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531–542.

Matthews, S. A. (2011). Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods. In L. M. Burton, S. P. Kemp, M. Leung, S. A. Matthews, & D. T. Takeuchi (Eds.), *Communities, neighborhoods, and health: Expanding the boundaries of place* (pp. 35–55). New York: Springer.

McLaughlin, D. K., & Stokes, C. S. (2002). Income inequality and mortality in US counties: Does minority racial concentration matter? *American Journal of Public Health, 92*(1), 99–104.

McLaughlin, D. K., Stokes, C. S., & Nonoyama, A. (2001). Residence and income inequality: Effects on mortality among US counties. *Rural Sociology, 66*(4), 579–598.

McLaughlin, D. K., Stokes, C. S., Smith, P. J., & Nonoyama, A. (2007). Differential mortality across the U.S.: The influence of place-based inequality. In L. M. Lobao, G. Hooks, & A. R. Tickamyer (Eds.), *The sociology of spatial inequality* (pp. 141–162). Albany: SUNY Press.

Miller, M. K., & Luloff, A. E. (1981). Who is rural? A typological approach to the examination of rurality. *Rural Sociology, 46*(4), 608–625.

Miller, M. K., Stokes, C. S., & Clifford, W. B. (1987). A comparison of the rural-urban mortality differentials for deaths from all causes, cardiovascular disease and cancer. *Journal of Rural Health, 3*(2), 23–34.

Morton, L. W. (2004). Spatial patterns of rural mortality. In N. Glasgow, L. W. Morton, & N. E. Johnson (Eds.), *Critical issues in rural health* (pp. 37–48). Ames: Blackwell Publishing Professional.

Mujahid, M. S., & Diez Roux, A. V. (2010). Neighborhood factors in health. In A. Steptoe (Ed.), *Handbook of behavioral medicine* (pp. 341–354). New York: Springer.

NCHS. (2011). *Compressed mortality file, 1999–2008* (Machine readable data file and documentation, CD-ROM series 20, No.2N). Hyattsville: National Center for Health Statistics.

Openshaw, S. (1983). *The modifiable areal unit problem* (Vol. 38). Norwich: Geo books.

Porter, J. R., & Howell, F. M. (2012). *Geographical sociology: Theoretical foundations and methodological applications in the sociology of location*. New York: Springer.

Preston, S., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Malden: Blackwell Publishers Inc.

Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Rupasingha, A., & Goetz, S. J. (2008). *US county-level social capital data, 1990–2005*. Accessed 17 Aug 2011.

Rupasingha, A., Goetz, S. J., & Freshwater, D. (2006). The production of social capital in US counties. *Journal of Socio-Economics, 35*(1), 83–101.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918–924.

Shoff, C., & Yang, T. C. (2013). Understanding maternal smoking during pregnancy: Does residential context matter? *Social Science & Medicine, 78*, 50–60.

Song, L., Son, J., & Lin, N. (2010). Social capital and health. In W. C. Cockerham (Ed.), *The New Blackwell companion to medical sociology* (pp. 184–210). Malden: Blackwell.

Sparks, P. J., & Sparks, C. S. (2010). An application of spatially autoregressive models to the study of US county mortality rates. *Population, Space and Place, 16*(6), 465–481.

Sparks, P. J., Sparks, C. S., & Campbell, J. J. A. (2013). An application of Bayesian spatial statistical methods to the study of racial and poverty segregation and infant mortality rates in the US. *Geojournal, 78*, 389–405.

Verheij, R. A. (1996). Explaining urban-rural variations in health: A review of interactions between individual and environment. *Social Science & Medicine, 42*(6), 923–935.

Voss, P. R. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*(5), 457.

Willits, F. K., Bealer, R. C., & Timbers, V. L. (1990). Popular images of 'rurality': Data from a Pennsylvania survey. *Rural Sociology, 55*(4), 559–578.

Wilmoth, J. M. (2010). Health trajectories among older movers. *Journal of Aging and Health, 22* (7), 862–881.

Yang, T. C., Chen, Y. J., Shoff, C., & Matthews, S. A. (2012). Using quantile regression to examine the effects of inequality across the mortality distribution in the US. *Social Science & Medicine, 74*(12), 1900–1910.

Yang, T. C., Jensen, L., & Haran, M. (2011). Social capital and human mortality: Explaining the rural paradox with county-level mortality data. *Rural Sociology, 76*(3), 347–374.

Yang, T. C., Noah, A. J., & Shoff, C. (2015). Exploring geographic variation in US mortality rates using a spatial Durbin approach. *Population, Space and Place, 21*(1), 18–37.

Chapter 14
Race, Place, and Space: Ecosocial Theory and Spatiotemporal Patterns of Pregnancy Outcomes

Michael R. Kramer

The social and spatial distributions of population health have interested social scientists and public health researchers for well over 100 years (Virchow 1848; Villerme 1830; Durkheim 1952). John Snow's geographic analysis of the mid-nineteenth century Cholera outbreak in London is perhaps the best known example of spatial epidemiology, but his attention to the social distribution of access to water sources exemplified the ways in which social begets spatial patterns in matters of population health (Johnson 2006). A detailed chronicling of the systematic relationships between race and class, place, and morbidity and mortality among African American's in late-nineteenth century Philadelphia gave rich context to how residential sorting and spatial distribution of resources produce social and spatial population health patterns (Du Bois 1899).

Epidemiology turned to a preoccupation with individual lifestyle and biomedical determinants of health in the mid to late twentieth century (Krieger 2011), but limited empirical and theoretical work persisted on the social allocation of health-relevant resources and exposures during this period (Cassel 1964; Syme and Berkman 1976; Yankauer 1950). The last 20 years have seen a dramatic increase in research on social determinants and socio-spatial patterning of health including a growing body of literature on neighborhoods and health (Kawachi and Berkman 2003; Acevedo-Garcia et al. 2008; Diez Roux and Mair 2010) and residential segregation and health (Acevedo-Garcia et al. 2003; Ellen 2000; Kramer and Hogue 2009a). The strong connection between social stratification and spatial allocation of economic opportunity lead Williams to term residential segregation a fundamental determinant of socio-spatial health disparities in the U.S. (Williams and Collins 2001).

M.R. Kramer (✉)
Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road, Atlanta, GA 30322, USA
e-mail: mkram02@emory.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_14

Despite the long history of—and recently renewed enthusiasm for—place and space as conduits of health-relevant distribution of exposures and resources, the theoretical framework for interpreting spatial or place-based patterns in health is incompletely developed, or not well ‘spatialized’. In this chapter I use the example of black-white racial differences in risk for a pregnancy outcome—low-birth weight-preterm birth—as an example for spatializing Krieger’s ecosocial theory and Geronimus’ life-course ‘weathering’ hypothesis. The chapter is organized in sections with a brief overview of biosocial determinants of preterm birth and low birth weight, followed by an introduction to the empirical population data to be used throughout the chapter, and then exploration and empirical results addressing three themes: spatial scale as a nuisance and a dimension for illumination, the life course accumulation of neighborhood experience, and temporal trajectories of places.

14.1 Ecosocial Theory and the Determinants of Pregnancy Outcomes

Preterm birth (infant born before 37 weeks gestation) and low birthweight (infant born less than 2500 g) are important population health indicators because of their association with infant mortality, long lasting morbidities including developmental and intellectual disabilities, and substantial associated medical and service economic costs (Callaghan et al. 2006; Bhutta et al. 2002; Behrman et al. 2007). Perinatal outcomes also represent a general model for conceptualizing and testing hypotheses about biological embodiment of social experience and the production of socially patterned health distributions (Kramer and Hogue 2009b; Hogue and Bremner 2005).

Large and persistent black-white racial disparities in infant mortality, low birthweight, and preterm birth have persisted into the twenty-first century despite dramatic improvements in prenatal and neonatal medical management (CDC 1999). Important individual-level risk markers for poor perinatal outcomes include maternal age, marital status, socioeconomic status, parity, substance abuse including smoking, genital tract infections including bacterial vaginosis, and shortened cervix (Goldenberg et al. 2008). Yet despite substantial effort to identify individual risk factors, these explain only a fraction of the black-white and socioeconomic gap (Kramer and Hogue 2009b).

The incomplete explanation of social inequity in pregnancy outcomes by biomedical and lifestyle perspectives has spurred development of a social epidemiology for perinatal health. Krieger summarizes three threads of social epidemiologic theorizing arising in the past several decades: sociopolitical, psychosocial, and ecosocial (2011). These are briefly reviewed, with attention to their application to perinatal outcomes.

The sociopolitical thread incorporates political economy, political power, relative position within social hierarchies and the social allocation of health opportunity (Link and Phelan 1995; Navarro and Muntaner 2004; Williams and Collins 2001). A range of sociopolitical hypotheses have been leveraged to understand racial disparities in pregnancy outcomes. DuBois (1899) and Yankauer (1950) noted in the early and mid-twentieth century respectively an increased infant mortality in areas characterized by economic and racial residential segregation, a finding which persists into this century (Ellen 2000; Kramer and Hogue 2009a; Kramer et al. 2010; Osypuk and Acevedo-Garcia 2008). Racial and economic residential segregation differentially allocate (or constrain) individuals to living and social environments characterized by uneven economic and educational opportunity, variable access to preventive health services, selective marketing by tobacco and alcohol interests, and multiple sources of social support and social stress (Williams and Collins 2001; Kramer and Hogue 2009a). This association between area-based residential segregation and individual perinatal outcomes may be mediated by area-level violent crime rates, as well as individual level smoking and prevalence of pre-conception chronic disease such as hypertension (Kramer et al. 2010; Grady and Ramirez 2008; Bell et al. 2007).

The psychosocial thread of epidemiologic theorizing emphasizes the mechanisms by which social experience is translated into biologic function or dysfunction. In this frame it is the individual biologic response to socially mediated experiences of relative inequality, discrimination, or stress which gives rise to differences in population health (Marmot 1988; Sapolsky 2004). Animal models and human studies suggest that exposure to acute and particularly chronic or repetitive socially-mediated stressors results in measurable and lasting changes in physiologic function including autonomic responses of blood pressure and heart rate, neuroendocrine profiles, and even anatomical differences in brain development if stress was experienced during critical developmental windows (McEwen 1998; Noble et al. 2005). The unique effects of chronic or repetitive stress on human health have been variously called allostatic load, weathering, and premature aging.

While possibly relevant for many chronic diseases, chronic stress is an increasingly accepted risk factor for preterm birth and low birth weight, and may be particularly important for understanding racial disparities in pregnancy outcomes (Behrman et al. 2007; Kramer et al. 2011). The stress-axis neuroendocrine transmitters such as cortisol, corticotropin releasing hormone and adrenocorticotropin hormone play a unique role in the healthy pregnancy, and women with a history of chronic stress have abnormal neuroendocrine profiles during pregnancy (Rich-Edwards and Grizzard 2005; Wadhwa et al. 2001). Thus women's stress experiences not just during pregnancy but in the months, years, or even decades before conception may prime her for a particular pregnancy trajectory (Halfon and Hochstein 2002).

The weathering hypothesis is a specific example of psychosocial theory (Geronimus 1996). To investigate whether exposure to chronic stress might prematurely age or 'weather' affected women, Geronimus noted that across human populations there is a consistent U- or J-shaped age-specific pattern for risk of

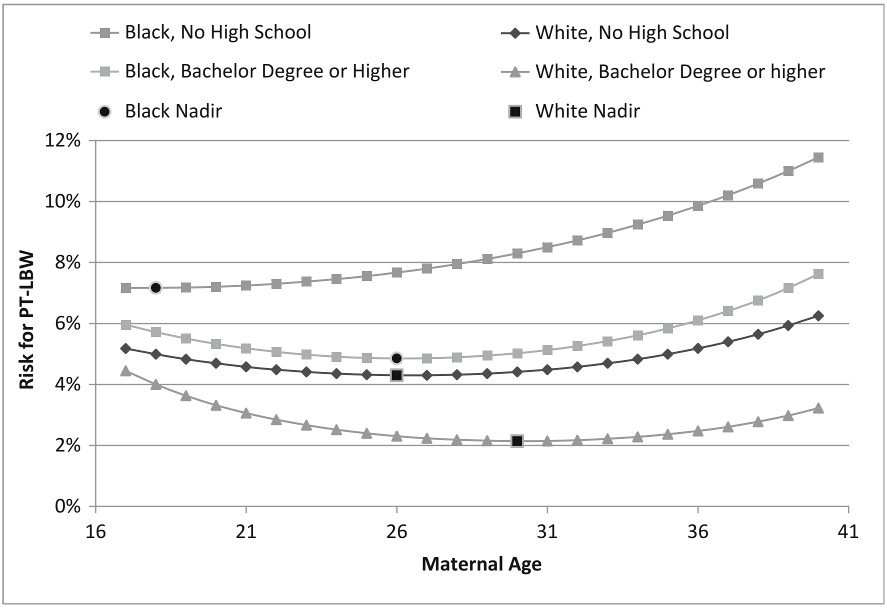

Fig. 14.1 Age-specific risk for LPTB stratified by race and attained education, United States, 2010. Legend: Age-specific risk for low birth weight-preterm birth (LPTB) varies by race and attained education. Younger nadir or minimal-risk age and steeper increase in age-specific risk is seen for black as compared with white women and for less educated as compared to more educated women. Data represents women in the 33 states which have adopted the 2003 revision of the U.S. standard birth certificate (76 % of all births in 2010) measuring education as attained degrees rather than years of education (Source of data: Martin et al. 2012)

preterm birth and low birthweight in which risk is highest among the youngest and the oldest mothers, with a nadir or optimal age at in the middle at which risk is lowest. When population risk is stratified by social dimensions, she noted that disadvantaged groups may demonstrate a younger nadir age (premature aging) and a steeper increase in age-specific risk (accelerated aging). For example in Fig. 14.1, the age-specific risk for preterm low birth weight among U.S. women in 2010 is stratified by maternal race and attained education. For both black and white women without a high school education, the nadir age is shifted to the left (younger age) as compared with same-race women with a college education. And for both levels of education, black women have both younger nadir ages, and steeper increase in the age-specific risk.

Nancy Krieger's ecosocial theory of disease distribution weds themes from both the socio-political and the psychosocial threads in an effort to understand the interrelation between the production and reproduction of social inequality as seen across levels of social hierarchy experienced throughout the life course (Krieger 2011, 2005; Krieger and Zierler 1997). Ecosocial theory conceives of the ecosystem in which population health is produced as defined by societal arrangement of power and resulting inequalities (e.g. by race/ethnicity, class, gender) as expressed at multiple spatial scales or levels from the individual to household to local, regional, or global, and at multiple temporal scales through history and across the life course. These notions of spatial and temporal scale and life course embodiment of social environment are directly relevant to the case of racial disparities in pregnancy outcomes and will be developed further in this chapter.

Thus ecosocial theory suggests that inequitable patterns of pregnancy outcomes within and between populations must be viewed in a spatio-temporally dynamic, relational, and multi-level framework. While some social epidemiologic research on pregnancy outcomes considers place-based characteristics, places such as neighborhoods or metropolitan areas are often viewed as fixed containers transmitting social experience in a cross-sectional manner. Thus there is a need to more completely spatialize extant theory for the social production of pregnancy outcomes, and to consider temporal dynamics in space/place characteristics, and temporal dynamics in women's social-spatial experience across the life course.

14.2 Spatializing Ecosocial Theory: Place, Race and Perinatal Health in Georgia

14.2.1 Individual Level Data

To further explore the spatiotemporal dynamics of racial disparities in pregnancy outcomes I use a population-based birth registry for all live birth to Georgia-resident women from 1994 to 2007. Composed of routinely collected birth certificate information, the Georgia maternally-linked longitudinal dataset is unique in two ways: the presence of maternal identifiers allowing the linkage of multiple births to the same woman (e.g. siblings), and the presence of high quality street-level geocodes for maternal residence at birth. The combination of these two features facilitates the creation of partial life-course residential trajectories and estimation of cumulative life course socio-environmental exposure (Kramer et al. 2013).

The pregnancy outcome of interest for this example is singleton live births who were born preterm (<37 weeks gestation) and low birth weight; this outcome will be called low birth weight-preterm birth (LPTB) for the remainder of the chapter. Black women as a group experience among the greatest excess risk of LPTB of any racial/ethnic group in the U.S., and because black and white women make up 86 % of all births in the study area, the focus here is on risk in non-Hispanic black and white women.

Because risk profiles and spatial patterns differ in metropolitan and non-metropolitan areas, for clarity, the examples discussed here focus on the 28-county Atlanta metropolitan statistical area (MSA) representing 54 % of live births in Georgia during the study period. Maternal residential address is routinely geocoded to the street level by the Office of Health Indicators for Planning of the

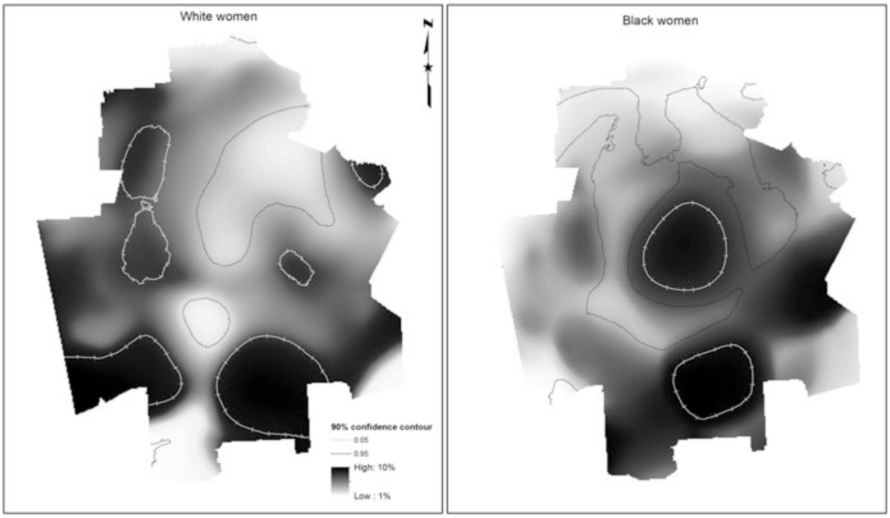

Fig. 14.2 Risk surface for LPTB among black and white women in Atlanta MSA, 2005–2007. Legend: Risk surfaces represent the spatially continuous variation in risk for LPTB as estimated with adaptive kernel smoothing of point data. 90 % Confidence Contours were estimated using random relabeling Monte Carlo simulation of the null distribution and represent areas where risk is significantly higher or lower than would be expected if risk were randomly distributed

Georgia Department of Public Health. Georgia vital records geocode quality has been assessed by comparing geocoded points to a 'gold standard' defined by tax parcel shapefiles and high resolution orthoimagery. The average spatial error in Atlanta was less than 100 m (Strickland et al. 2007). Only births with geocode quality at the street or census block level were included (7 % of observations excluded).

Overall LPTB risk from 1994 to 2007 for white women was 2.6 %, and for black women was 6.8 %. However this risk varies through time and through space. As seen in the maps of the risk surface for LPTB between 2005 and 2007 (Fig. 14.2) there is significant spatial variation both within and between racial groups. For black women, risk is significantly higher than average in the central MSA (roughly the City of Atlanta), but significantly lower than average in the suburbs surrounding the central city, as evidenced by the 90 % confidence interval contours. In contrast, for white women risk is overall lower, with pockets of elevated risk in southern and western counties.

Individual level covariates extracted from the birth certificate are considered because of their importance as correlates of pregnancy outcome (Table 14.1). As discussed above, unique patterns of risk occur by mother's age, and thus both age and age-squared are included as continuous variables. Individual socioeconomic status is hypothesized to be a marker for some forms of chronic stress related to poverty, and is correlated with restricted access to health services, health knowledge and information, and economic opportunity. Two measures of individual level

Table 14.1 Description of individual birth dataset for black and white women in the Atlanta MSA, 1994–2007

	White	Black
N[a]	236,385	141,700
LPTB (%)	2.6	6.8
Maternal age – mean (SD)	28.9 (5.6)	25.6 (6.1)
Maternal education (%)		
No high school	11.4	24.0
High school	22.4	35.5
Post-high school	66.3	40.5
Parity (%)		
First pregnancy	37.7	31.2
2nd or 3rd	55.1	53.1
4th or higher	7.1	15.7
Medicaid (%)	17.5	48.7
Married (%)	87.3	36.2
Smoke during pregnancy (%)	8.7	4.5
History of prior preterm birth (%)	12.6	17.1

LPTB low birth weight preterm birth (<2500 g and <37 weeks gestation), *MSA* metropolitan statistical area

[a]Analytic dataset is restricted to singleton, live births with gestational age between 20 and 44 weeks, and birthweight >500 g, with geocode match certainty at the street or census block level. Only births to women self-reporting as non-Hispanic white or non-Hispanic black are included

socioeconomic status are available. Attained maternal education (no high school, high school or GED, or post-secondary) at the time of delivery is a correlate of childhood and adult socioeconomic status. A second socioeocomic measure is whether Medicaid, a means tested public source of health insurance for low-income women, was the payor for delivery services (yes/no). Maternal smoking during pregnancy (yes/no) is a behavioral risk factor for LPTB, and risk also varies by parity (first pregnancy, 2nd or 3rd pregnancy, 4th or higher) and history of a prior preterm birth (yes/no). There was a national secular trend for increasing preterm birth during the study period; therefore year of birth is also captured.

14.2.2 Area-Level Measures

While commonly endorsed in health and place literature, the distinction between composition and context in neighborhood effects research is likely a false dichotomy. Observed spatial variation in pregnancy outcomes could occur from the intersection of spatiotemporally dynamic social processes, opportunity structures, and the residential and social selection of individuals with specific behaviors and exposures (Macintyre et al. 2002). Place attributes which may be particularly important for pregnancy outcomes include those related to social support, material and psychosocial sources of chronic stress, social disorder, and preventive health knowledge and opportunities (Morenoff 2003; Culhane and Elo 2005; Kramer and Hogue 2009b).

In this empirical case study, I use three measures of local spatial context aimed to tap into distinct hypothesized mechanisms. Each measure is calculated for each definition of 'local area' (see definitions of local neighborhood in next section), separately using 1990 and 2000 decennial census data, and 2005–2009 American Community Survey data. The first measure is the neighborhood deprivation index (NDI) developed from a multi-step theory-driven data reduction process and validated in multiple urban and suburban areas for pregnancy outcomes research (Messer et al. 2006, 2008; O'Campo et al. 2007). The NDI is a composite of five domains: poverty (% households below poverty line, % receiving public assistance, % with annual income <\$35,000 in 2007,[1] and % female headed households); occupation (% men employed in management); housing (% housing overcrowded); employment (% individuals over 16 who are in the labor force but unemployed); and education (% individuals over 25 without a high school education). The index ranges from −0.5 to +1.5 with higher numbers indicative of greater deprivation.

The second area-based measure is the index of concentration at the extremes (ICE), calculated as the [(# affluent households—# poor households)/total # households], where 'affluent' households are those with income of \$100,000 or more in 2007, and 'poor' households are those with income of \$35,000 or less[2] (Massey 2001). The index varies from −1 (all households are poor) to +1 (all households are affluent). While the NDI captures concentrated poverty and disadvantage, ICE highlights inequality in both tails of the income distribution and growing awareness of the importance of concentrated affluence as a conduit for access to protective social and institutional resources (Sampson et al. 2002; Brooks-Gunn et al. 1993; Pebley and Sastry 2003).

The third measure I term 'social disorganization', and use residential stability,[3] proportion of owner-occupied households, and proportion of female headed households as a coarse proxy (Jencks and Mayer 1990; Kubrin and Weitzer 2003). Residential instability, low rates of owner-occupied residency, and the concentrated poverty associated with high proportions of female headed households can indicate areas where barriers exist to strong social control and ties. The scale is a weighted average of the three components with weights proportional to principal components loading factors.

[1] Approximate inflation-adjusted cutpoints used for decennial census years: 1990: <\$25,000; 2000: <\$30,000.

[2] Cutpoints for decennial census were approximately inflation-adjusted to the 2007 values: 1990: <\$20,000 or >\$65,000; 2000: <\$30,000 or > \$85,000.

[3] Percent of households who moved in past 5-years as captured in the 1990 or 2000 decennial census. Percent of households who moved in past 1-year as captured in 2005–2009 ACS. Census Bureau working paper suggests that patterns of mobility are similar when comparing the 5-year definition from decennial census, and the 5-year pooled estimate of 1-year mobility from the ACS (Benetsky and Koerber 2012).

14.3 How Local Is Local? The Spatial Scale of Residential Environments

"[T]he local community is best thought of not as a single entity, but rather as a hierarchy of progressively more inclusive residential groupings. In this sense, we can think of neighborhoods as ecological units nested within successively larger communities" (Sampson et al. 2002; paraphrasing Suttles 1972). While geographers have looked extensively at issues of absolute and relative spatial scale (Meentemeyer 1989), far less attention has been paid to the meaning or inferential implications of scale in the study of spatial patterns in population health.

The vast majority of 'neighborhood effects' studies of health in the U.S. define neighborhood using the census tract, often—but not always—without even acknowledging the arbitrary nature of this choice. Researchers who do address the limitation of a single arbitrarily defined administrative definition of local place (e.g. census geographies) typically do so in terms of the statistical bias arising from the modifiable areal unit problem (MAUP). The MAUP concerns potential errors in quantitative measures and inference based on data aggregated to areas which are not conceptually linked to the underlying social process (Openshaw 1984). The bias arises from the arbitrary nature of the scale and zoning of aggregation. Reliance on census tracts in health research could be problematic if arbitrary changes in tract boundaries would result in different values of measured contexts or risks without real change in the underlying population, or if the process under investigation acts primarily at a larger or smaller scale than the tract (Lee et al. 2008).

However scale likely matters not just because of the statistical nuisance of the MAUP, but also in conceptualizing and testing the spatial nature of social processes in the first place. As Matthews and Yang (2013) point out, modern humans are spatially 'polygamous', not loyal to a single bounded place, but simultaneously occupying and experiencing multiple places for social, recreational, and economic purposes. A core proposition of ecosocial theory is that "determinants of current and changing societal patterns of disease distribution...are manifest at different levels [of power] and involve different spatiotemporal scales" (Krieger 2011, p. 215). An alternative to the arbitrarily bounded representation of place defined by census geography is a dynamic and 'relational' view of place, acknowledging the spatial 'polygamy' or heterogeneity of social experience at multiple spatiotemporal scales (Cummins et al. 2007). A relational view as compared to a conventional view of place might see space as multi-scale nodes in a network, with temporally dynamic place characteristics (e.g. 'declining' versus 'advancing'), and layers of assets contributing to opportunity structures (Cummins et al. 2007; Macintyre et al. 2002).

For example, research on segregation and health focusing solely on local neighborhood racial concentration as an indicator of segregation (e.g. Mason et al. 2009) may capture one aspect of the consequences of segregation. However the attributes (e.g. opportunity, composition, and human and social capital) of a neighborhood is not independent of the regional context in the presence of racialized housing and labor markets, or the state or national context of housing and development policy (Osypuk and Acevedo-Garcia 2010). Measuring segregation at different scales provides unique insight as compared to conventional methods reliant on a single arbitrary scale of census geography.

In proposing an explicitly spatial conceptualization and measurement approach to residential segregation, Reardon and colleagues contrast the segregation of metropolitan areas under alternate definitions of 'local' neighborhood, ranging from small walkable regions around one's home, to larger sub-regions of the city (Lee et al. 2008; Reardon and O'Sullivan 2004; Reardon et al. 2008). Not only do metropolitan segregation levels differ depending on the scale of local neighborhood, but the 'granularity' as expressed as the ratio of macro- to micro- patterns of segregation could be of substantive interest (Reardon et al. 2008). While theory may imply a particular spatial scale as optimal for a given question, conceptualizing the meaning of measure variance across scales may be worthwhile (Logan et al. 2010).

14.4 Comparing Bounded and Egocentric Neighborhoods in Metropolitan Atlanta

To illustrate and better understand the granularity or scale of neighborhood characteristics in the 28-county Atlanta metropolitan area, I operationalize two alternate classes of 'local neighborhood'. The first class of 'local' relies on census geography available: the block group, census tract, and county. These three census geographies are hierarchically nested (two or more block groups make up a tract, and tracts are sub-units of counties), and have been harmonized to 2000 boundaries for comparison. The median population size of block groups in the Atlanta MSA in 2000 was 1760 (inter-quartile range [IQR]: 1175, 2707) while median population in tracts was 6156 (IQR: 4005, 7604) and in counties was 73065 (IQR: 22871, 103641). The median area of block groups in Atlanta is 2.8 km^2 (IQR: 1.3, 8.5), while the median area for tracts is 8.7 km^2 (IQR: 4.1, 28.6) and for counties is 728 km^2 (IQR: 518, 1007).

The second class of 'local' definitions views space not as bounded areal units, but as a continuous surface. When area-based attributes are seen in this light, 'local' can be defined in an egocentric (Matthews 2011) manner where a neighborhood is the area surrounding a point in space. Specifically I adapt the conceptualization and operationalization of neighborhoods proposed by Reardon and colleagues (Lee et al. 2008; Reardon and O'Sullivan 2004) to create multi-scale egocentric neighborhoods for each birth in the Atlanta dataset and for each area-based measure.

The general approach begins with creation of an attribute surface representing the spatially varying density of persons or households of a given characteristic, such as 'black' or 'affluent'. This surface is created by converting census block group attribute data (the smallest unit for which demographic and economic variables are available across all three census') to a 100×100 m grid, with grid values expressing the density (e.g. persons or households per square kilometer) of the attribute. Each grid point represents the center of a series of overlapping but unique egocentric neighborhoods. The characteristics of each egocentric neighborhood are defined as a function of the values for all surrounding points. Specifically I utilize a bi-square kernel density function to define the neighborhood around a point as the weighted average of the surrounding points, with near points carrying more weight than far points. The bandwidth of the kernel function defines the extent or outer bounds of the egocentric neighborhood. Thus a 1 km bandwidth creates circular egocentric neighborhoods with a 1000 m radius. This process was repeated to create egocentric neighborhood surfaces for all area-based measures using four scales of local neighborhood: 1 km, 2 km, 4 km, and 8 km with constituent areas ranging from about 3–200 km^2. These scales span a range of activity spaces from the walkable micro-area surrounding around one's home to an area representing a sub-regions in a county where shopping, employment, worship, and recreation might take place (Lee et al. 2008; Sastry and Pebley 2010).

Thus there are seven definitions of neighborhood (block group, tract, county, and 1-, 2-, 4-, and 8-km egocentric) for which each of the area-based measures is calculated using 1990 and 2000 decennial census data and 2005–2009 5-year pooled American Community Survey data. The values for each location are linearly interpolated between census surveys to approximate the sociocontextual environment in which the mother lived in the year of each birth. Figure 14.3 contrasts the NDI, ICE, and social disorganization index under three definitions: census tract and 1- and 8-km egocentric. Figure 14.4 summarizes the distribution of women's residential NDI under each of the seven neighborhood definitions. For NDI (and for ICE and Social Disorganization, not shown here) there is generally greater variation in neighborhood environment for black women in Atlanta, with a minority of black women living in both the most and least deprived environments.

To test the association between neighborhood environment and risk for LPTB among black and white women in the Atlanta MSA, a series of spatial logistic regression models were fit with alternate definitions of neighborhood. The patterns of association vary by race, choice of area-based measure, and neighborhood scale (Fig. 14.5). Increasing deprivation (NDI) is positively associated with risk for LPTB, although the magnitude of association is stronger for white than for black women. The pattern is similar for ICE (although higher values of ICE represent salubrious not deprived environs).

The stronger relative magnitude of association for white as compared to black women may reflect racial differences in heterogeneity of both residential environments and risk for LPTB. The odds ratio modeled here is a relative measure of within-race association. Because black women experience twice the risk overall as do white women, a doubling of risk for whites represents a smaller absolute change than does a doubling of risk for blacks. Because the distribution of NDI for white women is narrower than for black women—meaning white women have more homogenous environments across the MSA than do black women—a 1-unit change

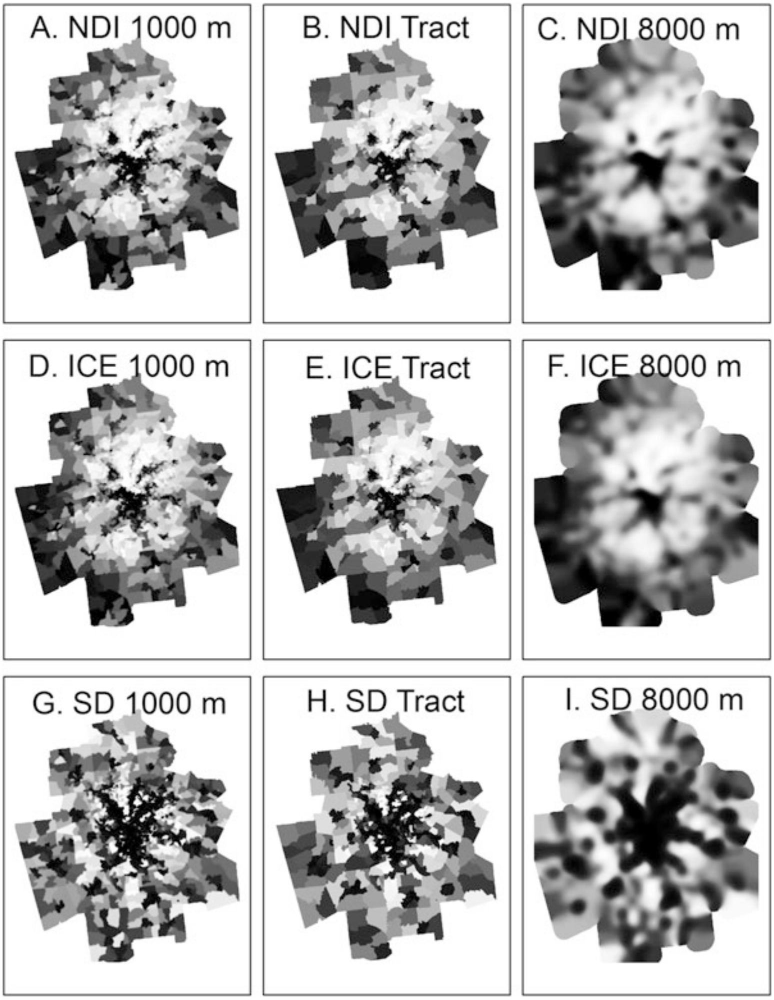

Fig. 14.3 Spatial distribution of area-based measures under three neighborhood definitions. Legend: Spatial variation in three 'neighborhood' areabased measures (along the rows: Neighborhood Deprivation Index [NDI], Index of Concentration at the Extremes [ICE], and Social Disorganization [SD]) as illustrated at three of seven measured spatial scales (down the columns: 1000-m egocentric, Census Tract, 8000- m egocentric). Dark shades represents lower opportunity environments (higher deprivation, more concentrated poverty, high social instability), while light shades represents greater opportunity environments (lower deprivation, concentrated affluence, greater social stability). (a) Neighborhood Deprivation Index, measured with 1000-m egocentric neighborhood. (b) Neighborhood Deprivation Index, measured at the Census Tract. (c) Neighborhood Deprivation Index, measured with 8000-m egocentric neighborhood. (d) Index of Concentrated Extremes, measured with 1000-m egocentric neighborhood. (e) Index of Concentrated Extremes, measured at the Census Tract. (f) Index of Concentrated Extremes, measured with 8000-m egocentric neighborhood. (g) Social Disorganization, measured with 1000-m egocentric neighborhood. (h) Social Disorganization, measured at the Census Tract. (i) Social Disorganization, measured with 8000-m egocentric neighborhood

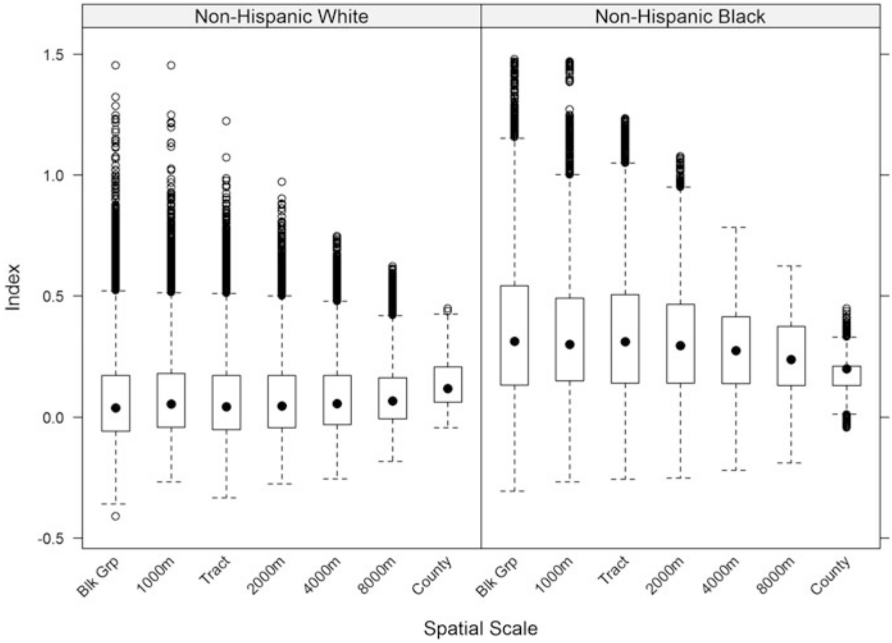

Fig. 14.4 Distribution of women's local Neighborhood Deprivation Index under seven definitions of 'local', Atlanta MSA, 1994–2007. Legend: *Box* and *whisker plots* of distribution of NDI for white and black women. *Black dots* represent the median, and the rectangles are the 25th and 75th percentiles (IQR). *Whiskers* represent the min/max of data or 1.5 times the IQR, whichever is less. *Circles* beyond the *whiskers* are outliers greater than 1.5*IQR from the rectangle. Distributions are similar for the Index of Concentration at the Extremes (ICE) and Social Disorganization (SD)

in NDI for white women encompasses a greater breadth of the within-race risk spectrum than does a 1-unit change in NDI for black women.

For NDI and ICE it also appears that the strength of association with LPTB for both white and black women increases with increasing neighborhood scale; it is not good to live in a micro neighborhood characterized by deprivation or concentrated poverty, but it is worse to live in an entire region or macro-area so characterized. This might support, for example, the importance of proximity to more affluent areas, and the social and economic opportunity structure which if not available in the immediate vicinity of one's home is at least available in the region.

The racial patterns seen for social disorganization—measured here as a composite of residential stability, tenure, and female-headed households—is somewhat different. At small spatial scales (block groups or 1- and 2-km egocentric neighborhoods) the association is stronger for whites than blacks. However as scale increases, the association weakens for whites, while strengthening for blacks. Considering the patterns of segregation and gentrification in Atlanta, this may point to group differences in how area-based measures are experienced. White and black women with higher area-based social disorganization micro-locally may be fully exposed to and

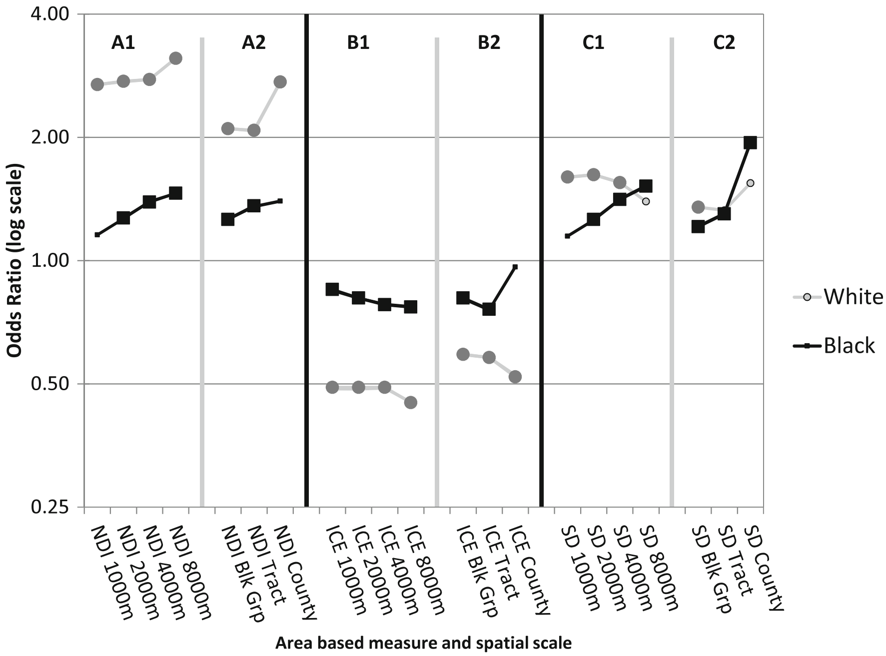

Fig. 14.5 Summary odds ratios for the association of area-based measures with low birth weight preterm birth, Atlanta MSA, 2005–2007. Legend: Odds ratios from spatial generalized additive (for egocentric neighborhood measures) and hierarchical logistic (for census areal unit neighborhood measures) regression, adjusting for maternal age, age-squared, year, smoking, parity, Medicaid, education, history of prior preterm birth and marital status. *Panel A* models association of Neighborhood Deprivation Index with LPTB (*A1* = egocentric neighborhoods, *A2* = census areal neighborhoods). *Panel B* and *C* report values for the Index of Concentration at the Extremes (ICE) and for Social Disorganization Index. Separate models were fit for black and white women. Confidence intervals excluded for clarity; however all odds ratios which are statistically significant at $\alpha = 0.05$ are represented with a large marker (*circle* for white women, *square* for black women) and odds ratios with p-value greater than 0.05 have small markers

part of this context. However white women living in broader regions of the MSA characterized by low residential stability and home-ownership may simply reside in gentrifying areas surrounded by more transitional neighborhoods.

While the general direction and significance of associations between area-based sociocontextual measures and women's risk for LPTB was consistent regardless of neighborhood definition, the method and spatial scale by which local environments are conceived and operationalized illuminates aspects of the underlying social process. The granularity of concentrated poverty or affluence, social instability, and deprivation could be indicative of higher level political and economic processes which shape and reproduce local places. An important shortcoming of this empirical example, then, is the constraint to study within a single MSA, rather than inclusion of areas across multiple MSA's or consideration of states or sub-national regions in which MSA's are situated. In other words expanding the scale of the ecosocial system even further would permit the comparison of patterns in MSA's racial and economic composition and segregation.

14.5 Dynamic Lives: Life Course Trajectories and Pregnancy Outcomes

Ecosocial theory highlights the multi-level and multi-scale interactions and relations in the social production of pregnancy outcomes. The preceding section emphasized the importance of spatial scale of neighborhood definitions and social processes, but these processes exist, vary, and operate across time as well as space. The growing interest in a life course perspective for understanding population patterns in pregnancy outcomes is due in part to sociologic and biologic evidence of the importance of early life environments—from an individual's own in utero environment to early childhood and adolescent family, school, and social context (Halfon and Hochstein 2002; Mishra et al. 2010; Collins et al. 2011). Evidence of trans-generational patterns in health may be due to non-heritable epigenetic gene expression as a result of socially-moderated environmental exposures, or from the inter-generational transfer of material wealth and cultural experience (Hogue and Bremner 2005; Collins et al. 2010).

To examine women's partial life course residential trajectories as predictors of subsequent pregnancy outcomes I build on previous work (Kramer et al. 2012, 2013) using the longitudinal linkages for births to Georgia women. In that work we demonstrated racial and socioeconomic differences in the life course trajectories for Georgia women, finding that cumulative more so than point-in-time measures were predictors of poor pregnancy outcomes. Using the Atlanta MSA births from 1994 to 2007, I now test Geronimus' weathering hypothesis of altered age-specific risk for LPTB as a function of cumulative social exposures. Specifically I test for differences in the nadir or minimum age for LPTB risk in sub-groups defined by race and cumulative spatially-measured NDI. The weathering hypothesis predicts that chronically stressed sub-groups will have a younger nadir as a result of premature aging of reproductively relevant neuroendocrine systems as compared to a less stressed population (Kramer et al. 2011; Geronimus 1996).

The construction and validation of the longitudinal linkages, and creation of the cumulative index of area-based measure has been described in detail elsewhere (Kramer et al. 2013). Briefly, women were observed across repeat pregnancies[4] and the range of neighborhood measures discussed in the previous section was captured for each pregnancy; for these analyses only women with two or more linked pregnancies are included. Due to space constraints, only NDI captured with the

[4] While the index or analysis pregnancies are solely for women residing in the 28-county Atlanta MSA at the time of delivery, the cumulative measure took account of all pregnancies in the state between 1994 and 2007.

2 km egocentric neighborhood is discussed here. Measures began at age 18 for all women. For women whose first birth was after age 18, we extrapolated the residential environment of their first birth back to age 18.[5] When the NDI differed at two points in time (e.g. two consecutive pregnancies), the values were linearly interpolated across the intervening years. The cumulative NDI index is defined as the sum of the NDI at each year of age from 18 to the year of the index pregnancy. Negative numbers represent the accumulation of exposure to low-deprivation neighborhoods, while numbers around zero or positive represent the accumulation of exposure to higher deprivation neighborhoods across the life course. In metro Atlanta the median cumulative NDI for black women was −0.3 (inter-quartile range: −4.4 to +3.0), while for white women the median was −10.3 (inter-quartile range: −16.4 to −3.4). In other words the racial differences in point-in-time NDI translate into a widening racial gap in cumulative NDI across the reproductive years.

Two approaches were taken to investigating the weathering hypothesis for pregnancy outcomes as it relates to spatially-varying social environments. The first was to fit a series of spatial generalized additive models separately for white and for black women where mean-centered age and age-squared are included with and without individual level covariates, and with and without interaction terms between cumulative NDI and age or age-squared. The intent of these models is to understand racial differences in the age-specific risk after accounting for possible variation in individual level factors, and modification by area-level factors. In the simplest model including only age and age-squared, and in subsequent more complex models with individual and area-level covariates, the nadir age for black women in general is younger than for white women (Table 14.2).

In the interaction models there is evidence that for black, but not white women, increasing cumulative NDI is associated with younger nadir, with the difference in nadir age of approximately 5 years comparing black women living at the 75th percentile of cumulative NDI as compared to women at the 25th percentile. This finding of interaction for black but not white women is consistent with previous findings (Love et al. 2010; Kramer et al. 2013). It is possible that this racial difference in the impact of life course cumulative neighborhood environment on pregnancy health is due to a threshold effect whereby few if any white women in population-based studies accrue the level of cumulative disadvantage necessary to make a measurable difference. On the other hand it is possible that even for a constant level of neighborhood deprivation there are racial differences in the meaning or experience, whether through the additive effects of racial as well as economic stratification and discrimination, or because of the presence of other unmeasured deleterious processes in black women or unmeasured buffering processes in white women.

[5] Sensitivity analyses were conducted excluding and/or controlling for an indicator for which women had measures extrapolated; final results were robust to these assumptions.

Table 14.2 Modeled nadir age[a] for LPTB among black and white women, Atlanta MSA, 1994–2007

	Nadir age	
	Black	White
Nadir age with no covariates	26.1	31.9
Nadir age with individual-level covariates[b]	26.9	33.3
p-value for interaction of Age x cumulative NDI	<0.0001	0.30
Nadir for women at 25th percentile of cumulative NDI[c]	29.4	NA
Nadir for women at 75th percentile of cumulative NDI[c]	24.6	NA

LPTB low birthweight, preterm birth, *NDI* neighborhood deprivation index, *MSA* metropolitan statistical area

[a]Nadir age is calculated from logistic regression coefficients using this formula: $(-\beta_{age})/(2^*\beta_{age\text{-}squared})$

[b]Adjusted for age, age-squared, parity, marital status, smoking, Medicaid, education, history of prior preterm birth, and year of birth

[c]Adjusted for all individual-level covariates plus NDI and the multiplicative interaction of NDI with both age and age-squared

A second approach to examining the variation in age-specific risk uses geographically-weighted regression (GWR) to test for non-stationarity of the association between age (and age-squared) and LPTB. Stationarity is the assumption that a relationship is constant through space; in the rubric of regression, stationarity is an assumption that a global beta coefficient is a valid summary measure of association between exposure and outcome (Kubrin and Weitzer 2003). Non-stationarity therefore is a violation of this assumption, or evidence that the relationship between an independent and dependent variable in a regression framework varies by, or depends on the location. For computational efficiency, and because spatial patterns can vary with time, this analysis is restricted solely to births from 2005 to 2007, thus demonstrating a pattern during only a portion of our overall study period.

Separate GWR models were fit for white and for black women. Tests for non-stationarity supported the use of a global, stationary relationship between age and risk for white women (Monte Carlo simulation p-value for test of non-stationarity $= 0.37$), but modest evidence for non-stationarity in the relationship between age and risk for black women ($p = 0.06$). Thus further GWR results are restricted to black women. The nadir age for black women at each location was calculated in the same manner as before, and is mapped in Fig. 14.6. Areas with the youngest nadir in age-specific risk for black women are the central portion of the MSA (roughly the city of Atlanta) with older nadir in age-specific risk in the north-central suburbs. This pattern is consistent with the spatial patterns of NDI generally, and also consistent with heterogeneity of black women's risk as a function of spatially-varying environments. What the GWR adds to this picture is the notion

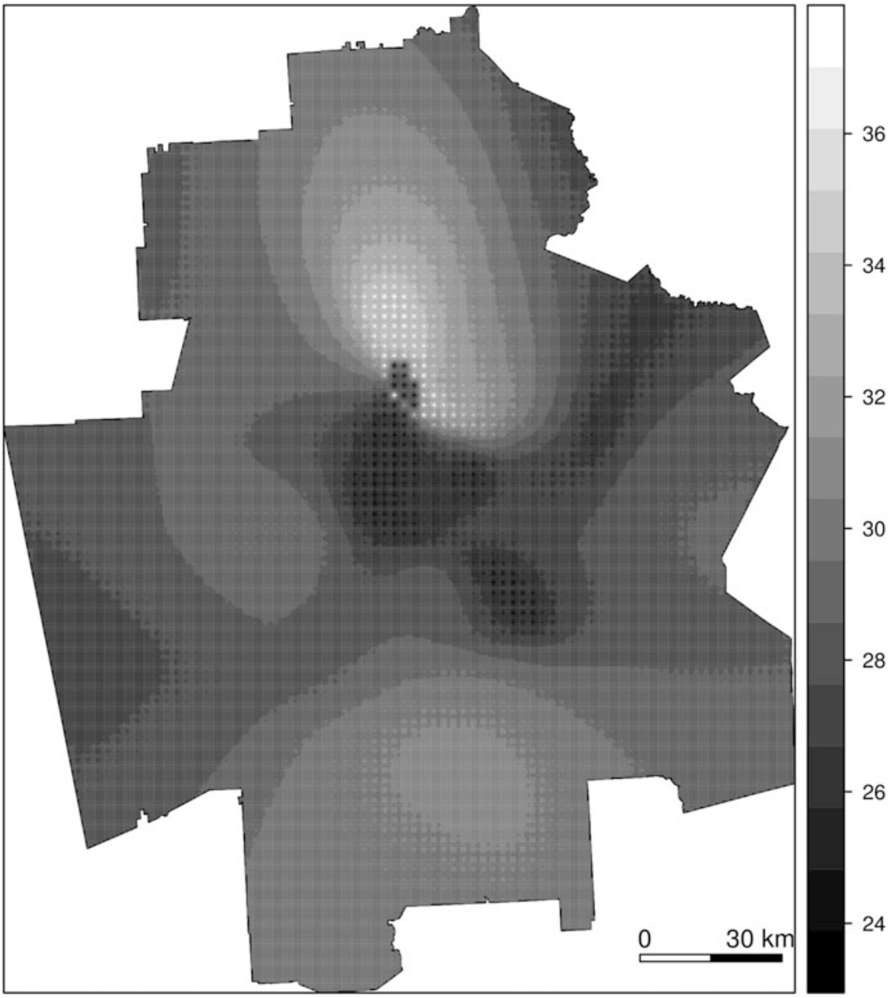

Fig. 14.6 Geographically weighted regression estimates of spatial variation of nadir in age-specific risk for LPTB among black women, Atlanta MSA, 2005–2007. Legend: Nadir in age-specific risk for low birth weight preterm birth (LPTB) calculated as $(-\beta_{age})/(2^{*}\beta_{age\text{-}squared})$ at each location in grid using spatially-varying regression coefficients from geographically weighted regression with adaptive bandwidth kernel estimator defining local weights

that not only is absolute risk varying, but also varying is an underlying pattern of risk by age which might previously have been presumed to be a biological constant, but appears to be modifiable by factors which correlate with area-based environment.

14.6 Dynamic Places: Neighborhood Trajectories and Pregnancy Outcomes

Ecosocial determinants of pregnancy outcomes likely accrue at specific critical developmental windows or through the accumulation of experience and exposure across the life course. Individuals experience changing social and contextual environments across the life course as a result of family and early life experience, economic and residential mobility, and interaction with opportunity structures. However places are also dynamic, traveling along trajectories with respect to the population composition, socioeconomic context, and embedded social processes (Cummins et al. 2007). In fact in residentially mobile populations, point-in-time measures of places likely underestimate the effect of context on individual and population health by misclassifying contextual experience (Murray et al. 2010). Two neighborhoods with similar point-in-time contextual measures may differ in meaningful ways if one is improving while the other is declining. These differences could impact the ongoing process of in and out migration, the optimism or pessimism of inhabitants about the future, and the range of likely investments of material and social resources from both inside and outside the neighborhood.

As a final empirical examination of spatiotemporal dynamics, I consider the association of neighborhood trajectory on LPTB risk. In other words I ask whether residence in ascending versus descending neighborhoods with respect to neighborhood deprivation, economic concentration, or social stability, is associated with pregnancy outcome holding constant the point-in-time neighborhood characteristics. Neighborhood trajectory is operationalized as the average annual absolute change in a given index value between two census survey points.

Annual neighborhoods trajectory ranged from negative to positive values for all measures at all scales for both races. For example the median index of concentration at the extremes measured with 1-km egocentric neighborhood for black women across the study period was −0.17. The median annual change for the ICE index among black women was −0.02 (IQR: −0.009, 0.002) meaning that there was a slight tendency towards increasing poverty concentration for the average black women in the study period. As with previous examples, this question is explored with the use of spatial generalized additive logistic regression models, fit separately for black and white women using all data from 1994 to 2007. In the interest of parsimony, only results from small (1 km egocentric) and large (8 km egocentric) neighborhoods are reported here.

There is a significant association between most point-in-time area-based measures and risk for LPTB across measures and scales with the exception of social disorganization for blacks (Table 14.3). While the direction of association between neighborhood trajectory and outcome is in the expected direction (e.g. neighborhoods on a trajectory of increasing or worsening NDI had greater risk of LPTB controlling for point-in-time measures) most trajectory associations were not statistically significant. Three exceptions which are borderline significant were the small-scale 1 km egocentric measure of the index of concentration at the

Table 14.3 Generalized additive model association of point-in-time and annual change in area-based measures with low birth weight preterm birth among black and white women, Atlanta, MSA, 1994–2007

		Black						White					
		1000 m			8000 m			1000 m			8000 m		
		OR	95 % CI		OR	95 % CI		OR	95 % CI		OR	95 % CI	
NDI	Point-in-time[a]	**1.19**	**1.08**	**1.30**	**1.43**	**1.22**	**1.67**	**1.71**	**1.43**	**2.04**	**2.07**	**1.61**	**2.65**
	Trajectory[b]	1.01	0.97	1.06	1.01	0.97	1.06	1.04	0.99	1.09	1.04	0.98	1.09
ICE	Point-in-time[a]	**0.81**	**0.75**	**0.88**	**0.72**	**0.63**	**0.81**	**0.71**	**0.64**	**0.79**	**0.64**	**0.56**	**0.74**
	Trajectory[b]	0.99	0.95	1.03	0.98	0.93	1.03	**0.93**	**0.88**	**0.99**	0.96	0.90	1.02
SOC	Point-in-time[a]	1.04	0.95	1.14	1.11	0.93	1.31	**1.30**	**1.13**	**1.50**	**1.32**	**1.06**	**1.66**
	Trajectory[b]	**1.05**	**1.00**	**1.09**	1.04	0.99	1.10	**1.06**	**1.01**	**1.13**	1.04	0.97	1.12

All models adjusted for age, age-squared, parity, marital status, smoking, Medicaid, education, history of prior preterm birth and year of birth. Bolded OR's and 95 % Confidence Intervals are statistically significant at $\alpha = 0.05$

NDI neighborhood deprivation index, *ICE* index of concentration at extremes, *SOC* social disorganization index, *OR* odds ratio, *CI* confidence interval

[a]Point-in-time estimates are the association between the area-based measure in the year of birth, and the risk of LPTB, conditional on individual covariates and trajectory

[b]Trajectory estimates are the association between the annual inter-censal change in the area-based measure and the risk of LPTB, conditional on individual level covariates and point-in-time area-based measures. Trajectories are standardized so that odds ratios represent the relative change in the odds of low birth weight preterm birth (LPTB) contrasting a neighborhood 1-sd above the mean trajectory to a neighborhood 1-sd below the mean trajectory

extremes for white women (trajectories towards increasing concentration of affluence is protective), and the small-scale egocentric measures of social disorganization for both black and white women (increasing social disorganization or instability is associated with increased risk).

In summary there is at best only modest evidence for an association between inter-censal change in area-based measures and risk for poor pregnancy outcomes. Such modest effects could be an accurate reflection of the relative importance of neighborhood dynamics above and beyond point-in-time context, or they could result from shortcomings in the means of measuring context and neighborhood change.

It should be noted that all analyses presented in this chapter make use of area-based measures which are linearly interpolated between census years, implicitly presuming that the change between two surveys was constant through time. This assumption may be particularly relevant in trajectory measures because it is unlikely that the 'rate of change' was equally experienced by all women in a given neighborhood in the inter-censal years. The study period encompasses a time of substantial economic and population growth in the Atlanta MSA, along with area-based changes such as urban gentrification, black suburbanization, and decommissioning of traditional public housing in an effort to deconcentrate poverty (Atlanta Regional Commission 2012; Kramer et al. 2012). In the case of public housing, the impacts of policy change on affected residents was not a gradual annual change across a decade, but occurred during a discrete period when a given housing project was decommissioned (Oakley et al. 2009). It is also challenging to model simultaneously the changing face of a place and the impetus for residential mobility in or out of the place. The complexity of such interactions is suggested by ecosocial theory, and is clearly of interest, but may need alternate analytic approaches ranging from geo-ethnography (Matthews 2011) to complex systems or agent based modeling (Entwisle 2007).

14.7 Conclusion

Theoretical perspectives for interrogating population patterns in health—and health disparities between socially-defined groups in particular—have a long history, but remain incompletely developed. Ecosocial theory and the complementary perspectives of the biological embodiment of multi-level interactions of power and social processes across the life course provide a rich framework for this work. The pronounced correlation between spatial patterns of residence and activity with the spatial distribution of health suggests that space and place is an important lens if not causal participant in the social production of health. But much work remains to be done to fully mine the ecosocial framework from a spatial point of view.

The three themes structuring this chapter are not the entirety of spatializing social epidemiologic theory, but they do point to the opportunity for work beyond the traditional static, cross-sectional, arbitrarily bounded health geography work which dominates the extant literature. Further attention to spatial scale not just as a statistical nuisance, but as an additional dimension for conceptualizing and measuring social environments and the interactions that occur within them can advance our understanding of place-health relationships. And just as a one-size-fits-all approach to scale is inadequate, so is a point-in-time approach to either individuals or places. Situating individual's health in the context of a life course trajectory of exposure and experience is consistent with growing evidence that the origins of chronic and reproductive disease may precede their occurrence by years, decades, and even generations. Finally, attention to the dynamic trajectories of places draws attention to the fact that social environments do not exist in a vacuum, but result from social and human capital investments of their inhabitants and from the larger processes of social stratification which contribute to spatially variable resource allocation (Entwisle 2007; Cummins et al. 2007).

References

Acevedo-Garcia, D., Lochner, K. A., Osypuk, T. L., & Subramanian, S. V. (2003). Future directions in residential segregation and health research: A multilevel approach. *American Journal of Public Health, 93*(2), 215–221.

Acevedo-Garcia, D., Osypuk, T. L., McArdle, N., & Williams, D. R. (2008). Toward a policy-relevant analysis of geographic and racial/ethnic disparities in child health. *Health Affairs (Millwood), 27*(2), 321–333.

Atlanta Regional Commission. (2012). *Regional snapshot*. http://www.atlantaregional.com/File%20Library/Info%20Center/Newsletters/Regional%20Snapshots/Population/RS_August_2012_Pop.pdf. Accessed 25 Apr 2013.

Behrman, R. E., & Butler, A. S. (Eds.). (2007). *Preterm birth: Causes, consequences, and prevention*. Washington, DC: National Academy Press, Institute of Medicine.

Bell, J. F., Zimmerman, F. J., Mayer, J. D., Almgren, G. R., & Huebner, C. E. (2007). Associations between residential segregation and smoking during pregnancy among urban African-American women. *Journal of Urban Health, 84*(3), 372–388.

Benetsky, M., & Koerber, W. (2012, May 3–5). *How do the ACS five-year migration data compare to the 2000 Census migration data?* Paper presented at the annual meeting of the Population Association of America, San Francisco, CA.

Bhutta, A. T., Cleves, M. A., Casey, P. H., & Cradock, M. M. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *JAMA: The Journal of the American Medical Association, 288*(6), 728–737.

Brooks-Gunn, J., Duncan, G., Kato, P., & Sealand, N. (1993). Do neighborhoods influence child and adolescent behavior? *American Journal of Sociology, 99*, 353–395.

Callaghan, W. M., MacDorman, M. F., Rasmussen, S. A., Qin, C., & Lackritz, E. M. (2006). The contribution of preterm birth to infant mortality rates in the United States. *Pediatrics, 118*(4), 1566–1573.

Cassel, J. (1964). Social science theory as a source of hypotheses in epidemiological research. *American Journal of Public Health, 54*, 1482–1488.

CDC. (1999). Achievements in public health, 1990–1999: Healthier mothers and babies. *MMWR. Morbidity and Mortality Weekly Report, 48*(38), 849–858.

Collins, J. W., Rankin, K. M., & David, R. J. (2010). Low birth weight across generations: The effect of economic environment. *Maternal and Child Health Journal, 15(4)*, 438–445.

Collins, J., Rankin, K. M., & David, R. J. (2011). African American women's lifetime upward economic mobility and preterm birth: The effect of fetal programming. *American Journal of Public Health, 101*(4), 714–719.
Culhane, J. F., & Elo, I. T. (2005). Neighborhood context and reproductive health. *American Journal of Obstetrics and Gynecology, 192*(5 Suppl), S22–S29.
Cummins, S., Curtis, S., Diez-Roux, A. V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Social Science and Medicine, 65*(9), 1825–1838.
Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences, 1186*, 125–145.
Du Bois, W. E. B. (1899). *The Philadelphia Negro: A social study*. Philadelphia: University of Pennsylvania Press.
Durkheim, E. (1952). *Suicide, a study in sociology* (International library of sociology and social reconstruction). London: Routledge & K. Paul.
Ellen, I. G. (2000). Is segregation bad for your health? The case of low birth weight. *Brookings-Wharton Papers on Urban Affairs, 2000*, 203–229.
Entwisle, B. (2007). Putting people into place. *Demography, 44*(4), 687–703.
Geronimus, A. T. (1996). Black/white differences in the relationship of maternal age to birthweight: A population-based test of the weathering hypothesis. *Social Science and Medicine, 42*(4), 589–597.
Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet, 371*(9606), 75–84.
Grady, S. C., & Ramirez, I. J. (2008). Mediating medical risk factors in the residential segregation and low birthweight relationship by race in New York City. *Health & Place, 14*(4), 661–677.
Halfon, N., & Hochstein, M. (2002). Life course health development: An integrated framework for developing health, policy, and research. *The Milbank Quarterly, 80*(3), 433–479.
Hogue, C. J., & Bremner, J. D. (2005). Stress model for research into preterm delivery among black women. *American Journal of Obstetrics and Gynecology, 192*(5 Suppl), S47–S55.
Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. E. Lynn & M. G. H. McGeary (Eds.), *Inner-city poverty in the United States* (pp. 111–186). Washington, DC: National Academy Press.
Johnson, S. (2006). *The ghost map: The story of London's most terrifying epidemic—and how it changed science, cities, and the modern world*. New York: Riverhead Books.
Kawachi, I., & Berkman, L. B. (Eds.). (2003). *Neighborhoods and health*. New York: Oxford University Press.
Kramer, M. R., & Hogue, C. R. (2009a). Is segregation bad for your health? *Epidemiologic Reviews, 31*, 178–194.
Kramer, M. R., & Hogue, C. R. (2009b). What causes racial disparities in very preterm birth? A biosocial perspective. *Epidemiologic Reviews, 31*, 84–98.
Kramer, M. R., Cooper, H. L., Drews-Botsch, C. D., Waller, L. A., & Hogue, C. R. (2010). Metropolitan isolation segregation and Black-White disparities in very preterm birth: A test of mediating pathways and variance explained. *Social Science & Medicine, 71*(12), 2108–2116.
Kramer, M. R., Hogue, C. J., Dunlop, A. L., & Menon, R. (2011). Preconceptional stress and racial disparities in preterm birth: An overview. *Acta Obstetricia et Gynecologica Scandinavica, 90*(12), 1307–1316.
Kramer, M. R., Waller, L. A., Dunlop, A. L., & Hogue, C. R. (2012). Housing transitions and low birthweight among low-income women: A longitudinal study of the perinatal consequences of changing public housing policy. *American Journal of Public Health, 102*(12), 2255–2261.
Kramer, M. R., Dunlop, A., & Hogue, C. (2013). Measuring women's cumulative neighborhood deprivation exposure using longitudinally linked vital records: A method for life course MCH research. *Maternal and Child Health Journal, 18*(2), 478–87.
Krieger, N. (2005). Embodiment: A conceptual glossary for epidemiology. *Journal of Epidemiology and Community Health, 59*(5), 350–355.

Krieger, N. (2011). *Epidemiology and the people's health: Theory and context*. New York: Oxford University Press.

Krieger, N., & Zierler, S. (1997). The need for epidemiologic theory. *Epidemiology, 8*(2), 212–214.

Kubrin, C. E., & Weitzer, R. (2003). New directions in social disorganization theory. *Journal of Research in Crime and Delinquency, 40*(4), 374–402.

Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review, 73*(5), 489–514.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behaviour, 35*(Special issue), 80–94.

Logan, J. R., Zhang, W., & Zu, H. (2010). Applying spatial thinking in social science research. *GeoJournal, 75*(1), 15–27.

Love, C., David, R. J., Rankin, K. M., & Collins, J. W. (2010). Exploring weathering: Effects of lifelong economic environment and maternal age on low birth weight, small for gestational age, and preterm birth in African-American and white women. *American Journal of Epidemiology, 172*(2), 127–134.

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and Medicine, 55*(1), 125–139.

Marmot, M. (1988). Psychosocial factors and cardiovascular disease: Epidemiological approaches. *European Heart Journal, 9*(6), 690–697.

Martin, J. A., Hamilton, B. E., Ventura, S. J., Osterman, M. J., Wilson, E. C., & Mathews, T. J. (2012). Births: Final data for 2010. *National Vital Statistics Report, 61*(1), 1–72.

Mason, S. M., Messer, L. C., Laraia, B. A., & Mendola, P. (2009). Segregation and preterm birth: The effects of neighborhood racial composition in North Carolina. *Health & Place, 15*(1), 1–9.

Massey, D. S. (2001). The prodigal paradigm returns: Ecology comes back to sociology. In A. Booth & A. Crouter (Eds.), *Does it take a village? Community effects on children, adolescents, and families*. Mahwah: Lawrence Erlbaum Associates.

Matthews, S. A. (2011). Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods. In *Communities, neighborhoods, and health* (pp. 35–55). New York: Springer.

Matthews, S. A., & Yang, T. C. (2013). Spatial polygamy and contextual exposures (SPACEs): Promoting activity space approaches in research on place and health. *American Behavioral Scientist, 57*(8), 1057–1081.

McEwen, B. S. (1998). Stress, adaptation, and disease. Allostasis and allostatic load. *Annals of the New York Academy of Sciences, 840*, 33–44.

Meentemeyer, V. (1989). Geographical perspectives of space, time, and scale. *Landscape Ecology, 3*(3–4), 163–173.

Messer, L. C., Laraia, B. A., Kaufman, J. S., Eyster, J., Holzman, C., Culhane, J., et al. (2006). The development of a standardized neighborhood deprivation index. *Journal of Urban Health, 83* (6), 1041–1062.

Messer, L. C., Vinikoor, L. C., Laraia, B. A., Kaufman, J. S., Eyster, J., Holzman, C., et al. (2008). Socioeconomic domains and associations with preterm birth. *Social Science & Medicine (1982), 67*(8), 1247–1257.

Mishra, G. D., Cooper, R., & Kuh, D. (2010). A life course approach to reproductive health: Theory and methods. *Maturitas, 65*(2), 92–97.

Morenoff, J. D. (2003). Neighborhood mechanisms and the spatial dynamics of birth weight. *American Journal of Sociology, 108*(5), 976–1017.

Murray, E. T., Diez Roux, A. V., Carnethon, M., Lutsey, P. L., Ni, H., & O'Meara, E. S. (2010). Trajectories of neighborhood poverty and associations with subclinical atherosclerosis and associated risk factors: The multi-ethnic study of atherosclerosis. *American Journal of Epidemiology, 171*(10), 1099–1108.

Navarro, V., & Muntaner, C. (Eds.). (2004). *Political and economic determinants of population health and wellbeing: Controversies and development*. Amityville: Baywood Publisher.

Noble, K. G., Tottenham, N., & Casey, B. J. (2005). Neuroscience perspectives on disparities in school readiness and cognitive achievement. *The Future of Children/Center for the Future of Children, the David and Lucile Packard Foundation, 15*(1), 71–89.

O'Campo, P., Burke, J. G., Culhane, J., Elo, I. T., Eyster, J., Holzman, C., et al. (2007). Neighborhood deprivation and preterm birth among non-hispanic black and white women in eight geographic areas in the united states. *American Journal of Epidemiology, 167*(2), 155–163.

Oakley, D., Ruel, E., & Reid, L. (2009). Moved out to where? Initial relocation destinations of former public housing residents in Atlanta a preliminary brief. Available via http://www.thecyberhood.net/documents/projects/pubho9.pdf. Accessed October 22, 2014.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography, 38*, 41.

Osypuk, T. L., & Acevedo-Garcia, D. (2008). Are racial disparities in preterm birth larger in hyper segregated areas? *American Journal of Epidemiology, 167*(11), 1295–1304.

Osypuk, T. L., & Acevedo-Garcia, D. (2010). Beyond individual neighborhoods: A geography of opportunity perspective for understanding racial/ethnic health disparities. *Health & Place, 16*(6), 1113–1123.

Pebley, A. R., & Sastry, N. (2003). Concentrated poverty vs. concentrated affluence: Effects on neighborhood social environments and children's outcomes (RAND Labor & Population Working Paper Series, pp. 03–24). Available at http://www.rand.org/content/dam/rand/pubs/drafts/2006/DRU2400.10.pdf. Accessed November 1, 2014.

Reardon, S. F., & O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology, 34*, 85–101.

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., et al. (2008). The geographic scale of metropolitan racial segregation. *Demography, 45*(3), 489–514.

Rich-Edwards, J. W., & Grizzard, T. A. (2005). Psychosocial stress and neuroendocrine mechanisms in preterm delivery. *American Journal of Obstetrics and Gynecology, 192*(5 Suppl), S30–S35.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*, 443–478.

Sapolsky, R. M. (2004). Social status and health in humans and other animals. *Annual Review of Anthropology, 33*, 393–418.

Sastry, N., & Pebley, A. R. (2010). Family and neighborhood sources of socioeconomic inequality in children's achievement. *Demography, 47*(3), 777–800.

Strickland, M. J., Siffel, C., Gardner, B. R., Berzen, A. K., & Correa, A. (2007). Quantifying geocode location error using GIS methods. *Environmental Health, 6*, 10.

Suttles, G. D. (1972). *The social construction of communities* (Studies of urban society). Chicago: University of Chicago Press.

Syme, S. L., & Berkman, L. F. (1976). Social class, susceptibility and sickness. *American Journal of Epidemiolgoy, 104*, 1–8.

Villerme, L. R. (1830). De la mortalite dans divers quarters de la ville de Paris. *Annales d'hygiene publique, 3*, 294–341.

Virchow, R. (1848). Report on the typhus epidemic in Upper Silesia. In L. J. Rather (Ed.), *Rudolph Virchow: Collected essays on public health and epidemiology* (pp. 205–220). Canton: Science History.

Wadhwa, P. D., Sandman, C. A., & Garite, T. J. (2001). The neurobiology of stress in human pregnancy: Implications for prematurity and development of the fetal central nervous system. *Progress in Brain Research, 133*, 131–142.

Williams, D. R., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports, 116*(5), 404–416.

Yankauer, A. (1950). The relationship of fetal and infant mortality to residential segregation: An inquiry into social epidemiology. *American Sociological Review, 15*(5), 644–648.

Chapter 15
Using Nighttime Lights Data as a Proxy in Social Scientific Research

Xi Chen

Given the shortcomings of standard sources of data, especially for countries with low-quality statistical systems or no national census or surveys, the possibility of using proxy measures for variables such as economic development, urbanization, population growth, and poverty holds great potential to help researchers dealing with mid-range issues. This chapter, divided into five sections, describes a methodological approach that uses geocoded, remote sensing information as a proxy measure for social scientific variables, and uses two applications to demonstrate how nighttime lights data can be used. The first section introduces two geocoded datasets, the satellite-based nighttime lights data produced by NOAA's National Geophysical Data Center (DMSP-OLS Nighttime Lights 2013) and the Geographically-based Economic Data (G-Econ 2013) produced at Yale University. The second section addresses theoretical and analytical issues of using lights data as a proxy for other variables in social scientific research. Using lights and GEcon datasets, the third section introduces a recent application of using lights as a proxy variable for economic statistics. Building off previous sections, the fourth section uses a lights-based proxy variable to test a theoretical proposition hypothesizing a negative effect of urbanization on poverty rates in developing countries. The final section presents conclusions drawn from the previous sections. The ultimate goal of this chapter is to introduce nighttime lights data and a formal statistical approach in using such data. The following sections demonstrate that lights data can be used as a proxy for many social scientific variables, especially those focusing on subnational areas and low-income countries.

X. Chen (✉)
Department of Sociology, Quinnipiac University, Hamden, CT, USA
e-mail: xi.chen@quinnipiac.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_15

15.1 Nighttime Lights and GEcon Data

15.1.1 Satellite-Based Nighttime Lights Data

Geocoded, remote sensing data in the form of nighttime lights, can be used as a proxy for variables commonly used in social scientific research. Such data holds great potential to provide social scientific information which in its traditional and direct form has been unavailable for certain regions of the world, or has been very costly to obtain. Furthermore, when data for these regions of the world is available, they are often of poor quality. The global satellite-based nighttime lights data, however, provides cheap, ready-to-use information for estimating many social scientific variables. Initially designed to collect worldwide low light imaging data of moonlit clouds, the nighttime lights data were recently developed from the Defense Meteorological Satellite Program- Operational Linescan System (DMSP-OLS 2013). The earliest light information was collected in the mid-1970s, and an archive for the digital data was established in the early nineties. The annual nighttime lights data from 1992 to 2010 are currently available with a resolution of 30 arc-sec[1] and covering 180°W to 180°E longitude and 75°N to 65°S latitude (DMSP-OLS 2013). Over the last two decades, the satellite-based nighttime lights data have been widely used by geographers and remote sensing scientists to measure social scientific variables in certain geographic areas (Elvidge et al. 1997, 2001, 2007; Sutton et al. 2007; Ebener et al. 2005). In these studies nighttime lights data was found useful to estimate socioeconomic variables at both national and subnational levels, including income per capita, wealth, and GDP (Ebener et al. 2005; Sutton et al. 2007; Elvidge et al. 1997; Noor et al. 2008). However, not until recently have other social scientists started paying attention to this valuable dataset (Henderson et al. 2012; Chen and Nordhaus 2011). Since lights data are available at a very high spatial resolution, it provides information for variables measuring human activities and impact of this activity in very small geographical areas (Sanderson et al. 2000). In addition to economic development and production activities, literature also suggests nighttime lights correlate with population density, urban development, and even CO2 emissions, and poverty rates (Henderson et al. 2012; Chen and Nordhaus 2011; Elvidge et al. 1997, 2009; Sutton et al. 2007; Ebener et al. 2005; Doll et al. 2000; Sanderson et al. 2000).

Figure 15.1 represents the image of nighttime lights for a large part of Asia, Europe, and Africa. From the picture it is obvious that Europe, East China, and India have extensive areas of bright lights, while in contrast, the majority of Africa is dark except for areas along the north Africa coastal line and in Africa's largest cities. The bright lights shining along the Nile and its delta region in Egypt illustrate

[1] Instead of the decimal degree (DD) coordinate system, the degree-minute-second (DMS) coordinate system is primarily used in this chapter. The geocoded data sources, including lights and GEcon, are introduced and presented with DMS system at their source websites. Using consistent unit with those presented in raw data can help researchers download and use these data.

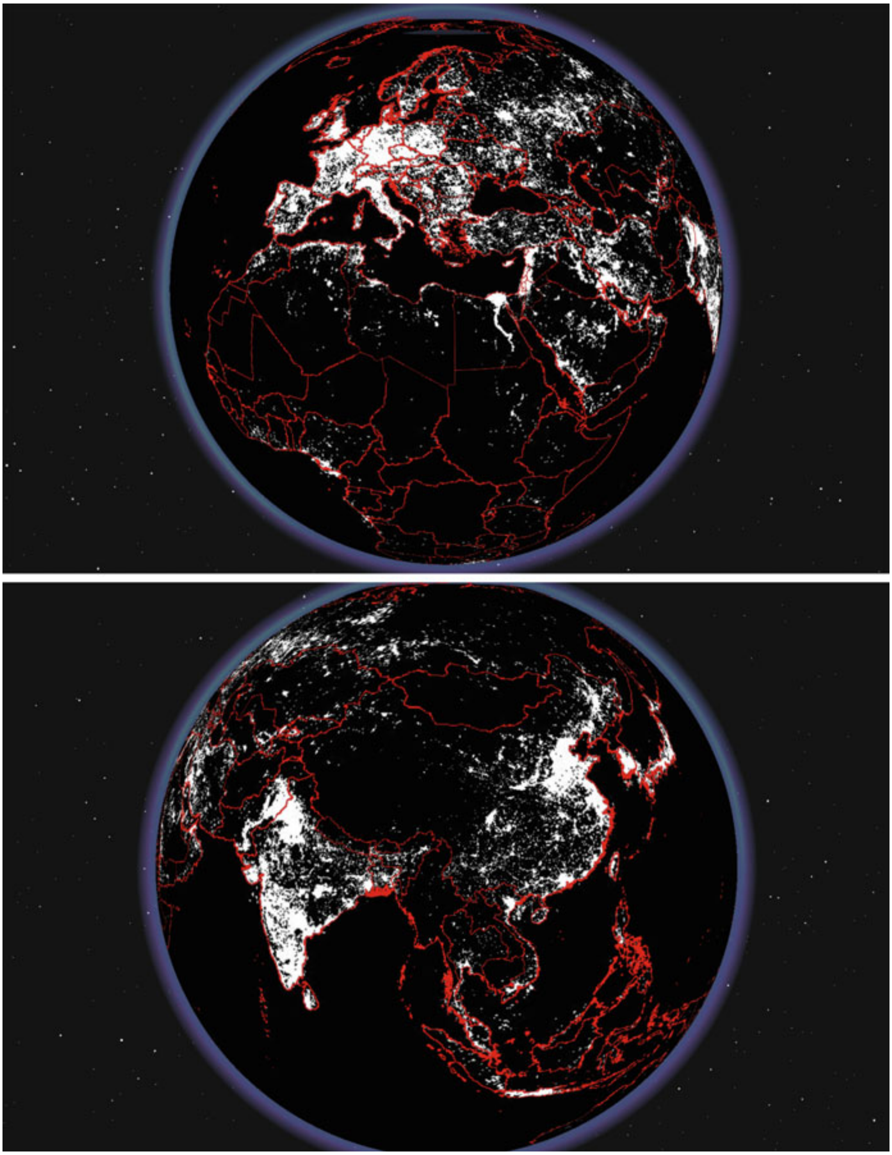

Fig. 15.1 The global image of nighttime lights (stable lights for year 2006)

the possibility of using high resolution image to distinguish socioeconomic differences across subnational areas. Furthermore, because the lights information is collected within the same time frame, at the same resolution at the global scale, it is a valuable source for researchers who focus on cross-national comparison and global patterns. Utilizing such a data source may help overcome a lack of standard

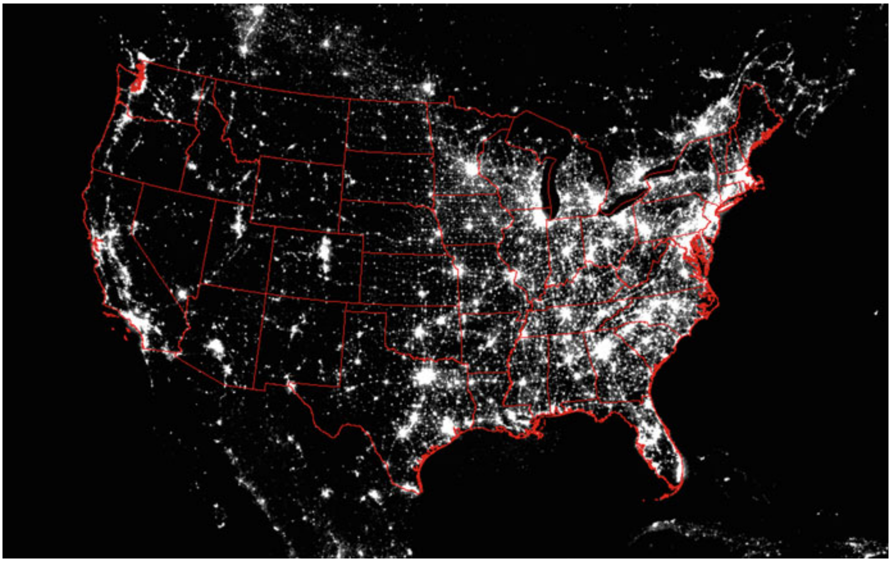

Fig. 15.2 US contiguous map of nighttime lights (stable lights of year 2006)

measures in less-developed countries, but also, may be beneficial for studies on more developed countries where high quality conventional statistical data are available, as proxy light measures may provide *other* information that researchers have not attended to. In a contiguous map of the US, as shown in Fig. 15.2, one can easily identify large cities, suburban areas, small towns, unoccupied land, and even interstate highways. Their temporal and spatial development and patterns can be examined through the dynamics in light intensity over time and across pixels. In sum, compared to data collected from conventional surveys or census by various statistical agencies, the nighttime lights data is "objectively" measured, updated instantly and regularly, and is universally measurable except for at high latitudes. NOAA's National Geophysical Data Center processes the raw lights into an annual cloud-free light dataset and provides open access to the time series. It is undoubtedly a low-cost and readily accessible source of data that can potentially advance research in the social sciences.

There are multiple versions of nighttime lights data currently available from DMSP-OLS: the "raw," the "stable," the "calibrated" and newly generated VIIRS nighttime lights. The stable annual lights data is processed from the raw lights, available from 1992 to 2010, and is the data product most often used in current literature. The radiance calibrated lights data has many advanced features over stable lights but is only available for 2006. Detailed methods for producing the stable and radiance lights are summarized by Baugh et al. (2010) and Ziskin et al. (2010). The global cloud-free VIIRS lights data is the most recent product from NOAA's National Geophysical Data Center: It spans 65° South to 75° North latitude, and is in 15-arc sec, making it the dataset with the finest resolution among

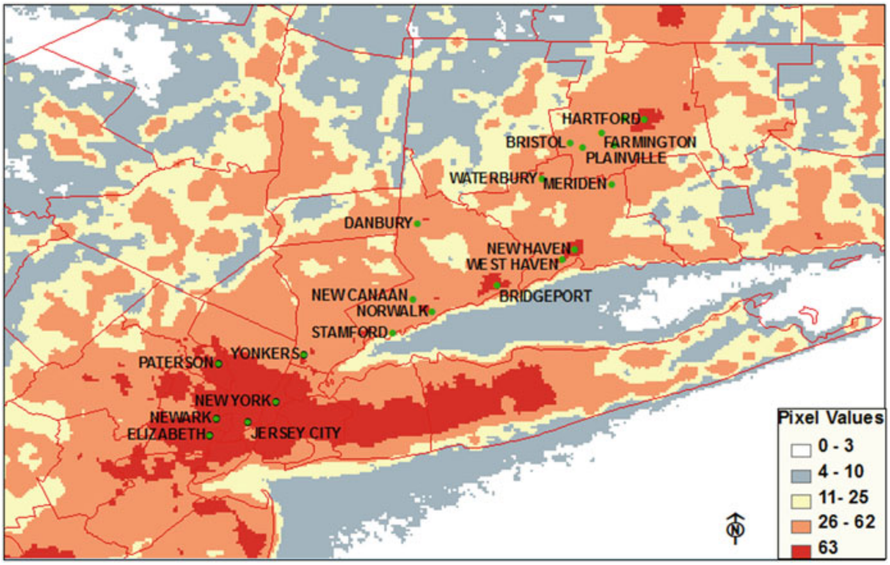

Fig. 15.3 Map of Long Island, New York regions with color-coded pixel values (stable lights for year 2006)

all lights products. However, the VIIRS is only available for the year 2012 and has not been cleaned up to remove aurora, fires, flares and sky-glow reflections off of snow, ice, and dry lake beds.

Both raw and stable lights data are presented as digital numbers (DN) ranging from 0 to 63 for each pixel. The stable lights data is processed from the raw lights with complex procedures to remove background noise (Elvidge 2001). Few studies have tested whether using raw lights compared to stable lights leads to different results in the analysis. In examining economic output and lights at the national and grid cell level, Chen and Nordhaus (2011) used sensitivity analysis to explore this issue and found only a small quantitative difference using the different light products.

Although, stable lights data is the most often used, it has serious problems of saturation due to top-coding at 63 and over-glowing across pixels. Figure 15.3 illustrates examples of such problems. The massive region in red indicates light saturation around the New York City metropolitan and surrounding areas. A large number of pixels in these areas are coded as 63, and this top-coding obviously cannot distinguish the difference between Elizabeth, New Jersey and lower Manhattan areas in and around New York City. On the other hand, the gray color of the water surrounding Long Island shows over-glowing problems. The DNs of pixels in these areas range from 4 to 10 instead of 0 because bright lights from the city over glow due to the presence of reflection from nearby water. In contrast the DN of calibrated lights ranges from 0 to 6030, which has less saturation problems at the highest intensities, and less over glowing problems. A side-by-side comparison of the stable and calibrated lights in Fig. 15.4a, b further illustrates the differences.

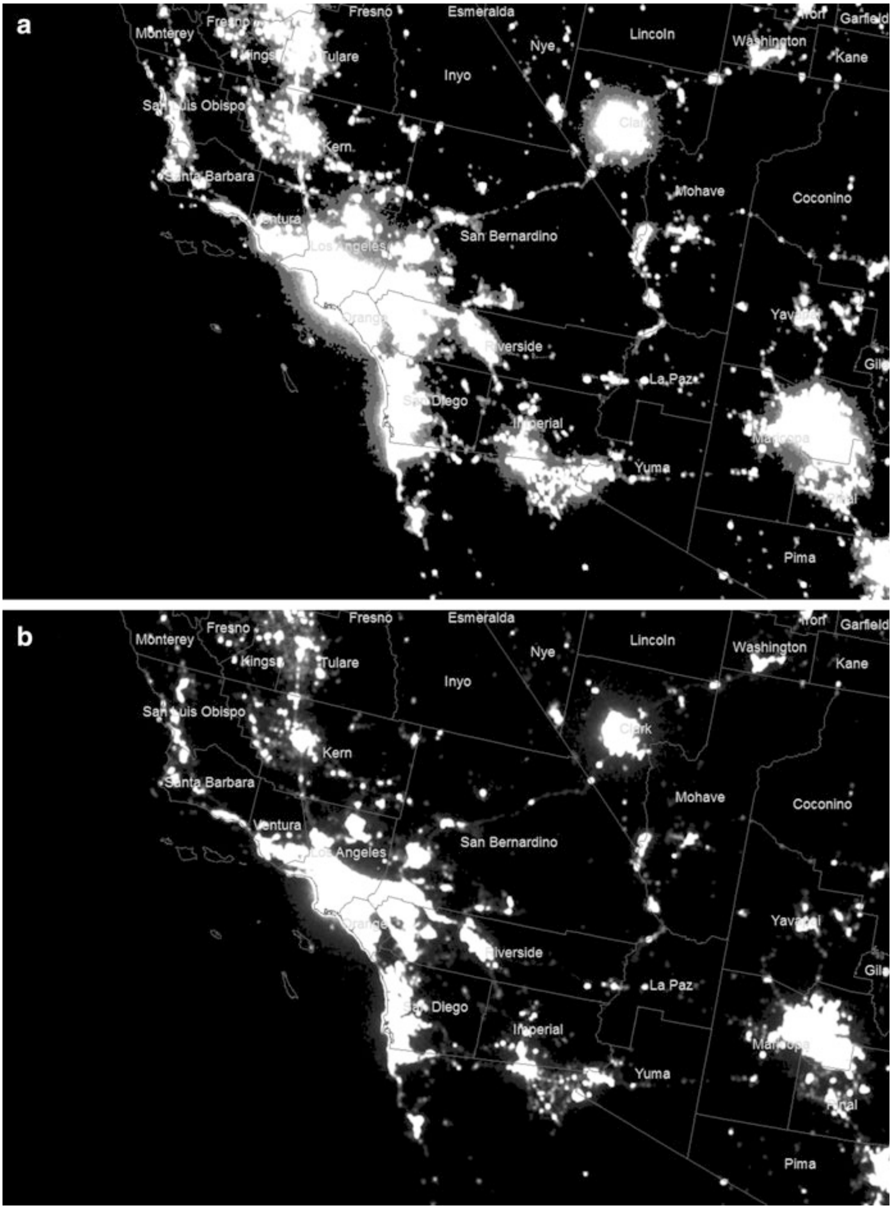

Fig. 15.4 (**a**) Los Angeles and its coastal areas (stable lights for year 2006; digital number ranges from 0 to 63) (**b**) Los Angeles and its coastal areas (calibrated lights for year 2006, digital number ranges from 0 to 6030.77)

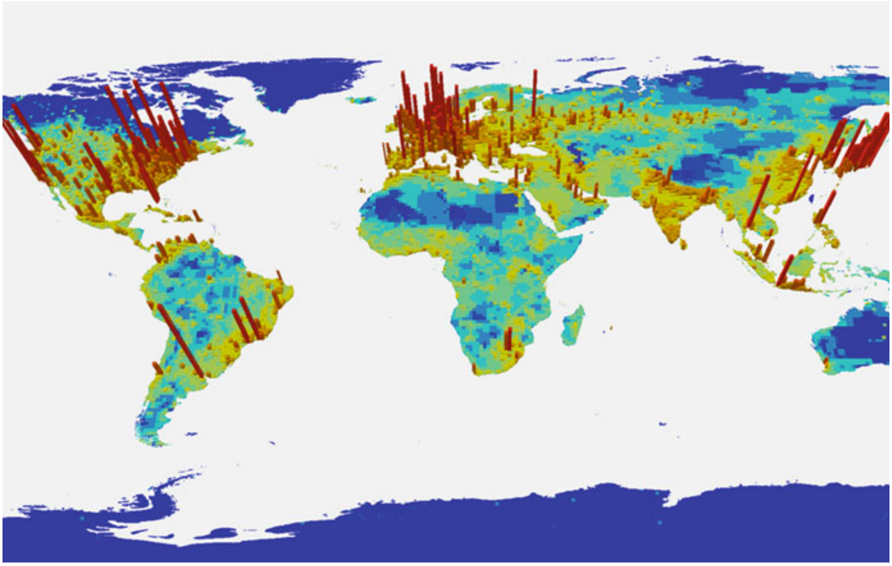

Fig. 15.5 3D image for grid cell output (GCP) data for year 1990 (The raw data is obtained from geon.yale.edu)

15.1.2 The GEcon Data

Another important source of geocoded data used in nighttime lights proxy research is the GEcon dataset. This dataset is housed by Yale's Department of Economics and its official website, gecon.yale.edu, provides global and individual country records and detailed descriptions on data collection and spatial allocation methods used to generate a final data product. Organized around geophysical boundaries of 1-arc degree by 1-arc degree, the GEcon data for the year 1990 was first published in 2006. Over the years, the dataset has been updated with more observations, improved methods, and now includes the years 1990, 1995, 2000, and 2005.

The Yale GEcon data primarily are economic output accounts for 1-arc degree grid cells, gross cell product, or GCP. Such units divide the surface of the entire globe into 64,800 cells, with most of these cells demarking oceans. The GEcon dataset provides 27,442 observations of terrestrial surfaces of the globe, including cell information not only on economic output, population, land area size, but also on basic climate measures, such as precipitation, temperature, and geophysical measures, such as soil, vegetation categories and distance to navigable rivers and ice-free oceans. The 3D image illustrates the grid cell economic output for the year 1990 at the global scale (Fig. 15.5), in which the cells are extruded according to the quantities of economic output they generated. Concurring with regional development patterns, the taller bars are more heavily concentrated in the US, Europe, Japan, costal areas of China, and large cities in other countries.

The grid cell measures have several advantages over information collected according to political administrative boundaries (Nordhaus et al. 2006). First, high-resolution geophysical units, e.g. one arc-degree or minute, can present information at multiple scales, and some can be much smaller than what has been traditionally defined by national or level I administrative boundaries for large countries. Note that the land area of one arc-degree cell is equivalent to 10,000 km^2 around the equator, but shrinks in the east-west direction when moving away from the equator toward north and south poles. Thus GEcon data provides economic and other estimates at a much smaller scale than standard national accounts. Second, geocoded social scientific variables, e.g. grid cell population or economic output, can also be conveniently merged with other geophysical information such as latitude, temperature and precipitation, and therefore allows for geophysical considerations in social science research.

Additionally, the Gridded Population of the World (GPW) and Global Rural-Urban Mapping Project (GRUMP) are other important sources of geocoded data for social scientific research. Over the last decade, demographers and geographers have linked and utilized the spatial characteristics of such data to integrate attributes capturing the physical and social world in a wide range of population and environment studies, such as in investigation urban settlements and the risks of climate change (Balk et al. 2009), ecosystem conditions and human well-being (DeFries et al. 2005), and infectious disease modeling (Tatem et al. 2012).

In the GEcon data, the main variable is GCP, which is measured in a similar fashion as gross domestic product (GDP) and gross regional product, except that the unit of the measure is 1-arc degree. The formal definition of GDP is the total production of market goods and services in a region less purchases from businesses (Nordhaus et al. 2006). Since there are no official GDP records at the grid cell level, a spatial rescaling approach is used in estimating GCP. The following section on the description of spatial rescaling procedure is partially taken from data documentation from the Yale GEcon website (Nordhaus et al. 2006). For most countries, GDP or per capita GDP are only available by political boundaries. Spatial rescaling, also known as "cross-area aggregation" or as "areal interpolation," can transform national or subnational accounts to spatially-based accounts. Figure 15.6 provides a visual demonstration of such an approach used in producing the GEcon data. In the figure, assuming A and B are two adjacent subnational areas, e.g. states that cover 12 grid cells, we can interpolate GCP values for 12 grid cells based on available state-level economic and demographic data. Currently, there are multiple approaches to spatially disaggregating information, including seven techniques investigated by Nordhaus and his colleague (2006): "(1) weighted average or proportional allocation, (2) median allocation or plurality rule, (3) local kernel regression (six alternatives), (4) global kernel regression (three alternatives), (5) weighted non-linear regression, (6) country average, and (7) pycnophylactic smoothing," but only the proportional allocation technique is used to produce the GEcon data. With this technique, the grid cells are first divided into "sub-grid" cells, which belong to the smallest available political administrative unit. For example, in Fig. 15.6, the grey-highlighted cell is divided into two subcells by

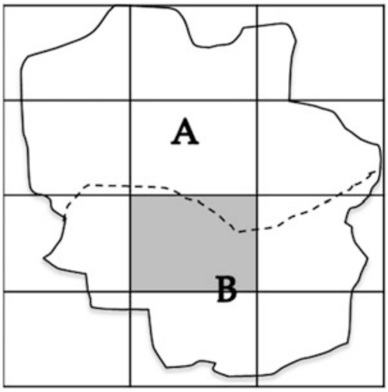

Fig. 15.6 Illustration of spatial allocation technique used in constructing the GEcon data

the borderline (dotted line) between state A and B. For a grid cell that falls completely within a state, the GCP is calculated as the multiplication of the grid cell population and GDP per capita of the state. Here, the grid cell population account is provided by The Gridded Population of the World (GPW) estimates for all terrestrial grid cells for the years 1990, 1995, 2000 and 2005 (CIESIN 2013), and state GDP per capita can be obtained from standard state accounts published by national statistics offices. For the calculation of subcell accounts, population density in each state is assumed to be distributed uniformly, and therefore a subcell population can be allocated according to the share of the subcell area to the state area and state population density. In other words, the population of two subcells in gray are estimated according to the area size of each subcell and population density for state A and B, respectively. Next, the subcell population is rescaled to confirm the GPW estimate of the total population of grid cells, and then the rescaled subcell populations are multiplied with the respective state per capita output or per capita GDP to calculate subcell economic output. Again, per capita output is assumed to be uniformly distributed in each state. The gross cell product (GCP) is the sum of output from all subcells that are located in the grid cell. Finally, the total GCP of a country is rescaled to conform to the total national GDP published by the World Bank in purchasing-power-corrected 2005 international U.S. dollars.

Not all countries are estimated with the same level of demographic and economic data. Depending upon the data availability and quality of each country, there are different levels of data which can be drawn upon in generating grid cell product (Nordhaus et al. 2006). For most small countries, defined as a country with less than 50 grid cells, such as Ghana, Lithuania, and North Korea, the national economic output data are used in the spatial rescaling process. National per capita output is assumed constant across all grid cells within a country. For such countries, the distribution of GCP is only influenced by population distribution. For countries that have regional accounts developed by national statistics agencies, either state/province data measured at the first political subdivision level or county/municipality data measured at the second political subdivision are used in spatial rescaling. For instance, for Australia, Canada, India, and Mexico, data at the first political subdivision are used, while for US, China, Brazil, and South Africa, and European Union countries, data collected at the second political subdivision level are used. Because most subnational economic data were provided by national statistics agencies, the quality of data varies widely by country and cannot be precisely estimated. In general, for high-income countries that have standard subnational statistical accounts, the constructed GCP estimate is more reliable. However, for low-income counties that often do not have official national statistics agencies or regular surveys or census, their GCP estimates are much less reliable (Nordhaus et al. 2006).

Already, the GEcon data has allowed for consideration of geospatial factors in investigations of social scientific questions, such as those looking at the relationship between temperature and grid cell output, the impact of geographic attributes on African poverty, and economic damages caused by greenhouse warming (Nordhaus 2006). Researchers have also utilized this dataset in studies on other economic and environmental topics (Buhaug et al. 2011; Tang and Woods 2008; Seo 2011), and even on inequalities and ethnic conflict (Cederman et al. 2011). The 1×1 unit estimates of economic output allow researchers to investigate human activities and their impacts occurring in much smaller geographic areas than conventional national or subnational political division, and to merge geocoded economic data with other geophysical information to test the relationship between such variables. Because of this advantage, the GEcon data can be easily merged with lights data at the grid cell level, and used to test the amount of information that is provided by lights as a proxy measure. With these fist steps now established, the next two sections will discuss analytic models and recent findings in literature focusing on lights and economic statistics.

15.2 Analytical Background

A National Research Council report recently emphasized formal statistical approaches in using proxy measures (North et al. 2006). This section provides detailed explanation on the analytical framework in Chen and Nordhaus (2011) and Henderson et al. (2012), two recent studies using an error-measurement approach in evaluation of lights data as a proxy for economic statistics.

Although the chapter mainly uses nighttime lights data to demonstrate the method proposed, other geocoded information can also be considered as a proxy using the same method for studying social scientific questions. Utilizing a wide range of geocoded or remote sensing data as proxies bypasses problems faced with poor quality, time restricted and expensive conventional survey methods, encountered with traditional data collection attempts from some parts of the world. Using available geocoded remote sensing information potentially improves current estimation of such existing data. However, before proxies can be utilized, the main question which needs to be addressed is exactly how much information such proxy variables can provide in improving the estimation of traditionally-collected variables? In other words, how can we use a statistical model to test the usefulness of proxy variables?

The analytic models provided by both Chen and Nordhaus (2011) and Henderson et al. (2012) answer the above question. Both studies present very similar statistical approaches to test how much information nighttime lights can provide to estimate economic statistics. Both articles assume that there exists true real economic output, and that the observed values of such variables are subject to measurement errors. They also assume that there is a structural econometric relationship between the light measures and economic output variables, and therefore lights can be used as a proxy measure to improve our estimation of output variables. Finally, they assume that the error term in the structural relationship between output variables and the lights variable is primarily caused by measurement errors in lights data.

Such errors in the measures of lights are due to multiple sources (Chen and Nordhaus 2011). It is due in part to the fact that lights data are collected by different satellites. In addition, degrading optical qualities of a satellite over the years can also cause measurement error. Finally, lights data is collected at night, while most human activity, including economic production, is conducted mainly during the daytime for most locations on the globe. Therefore, given a causal relationship between economic development and lights intensity, the error term in the structural equation, or mean squared residual, are primarily from errors in measuring lights.

While both articles agree on the main assumptions, the identification process of the estimation of optimal weights of lights is different. The identification in Nordhaus and Chen (2015) relies on estimates of the measurement error of conventional economic output measures, while the later study relies on a signal-to-noise ratio between lights and income to derive the optimal weights of lights (Henderson et al. 2012). Regardless of this difference, both articles apply optimal weighting in combining proxy and conventional variables, which allows the authors to assess the usefulness of lights data as a proxy variable.

Here we will only address the analytical model proposed by Chen and Nordhaus (2011), with all formulas taken from Supporting Information in their published article. To answer how much information proxy variables can provide in improving traditional measures, the first step is to use mathematical language to express the assumptions and relationships between proxy variables and the variables that we intend to study. Let us assume the variable of our interest, Y, is economic output or GDP. We would like to know how much information a proxy variable, such as nighttime lights, can provide to improve our measure of GDP. We can define:

y = measured or observed value of GDP
y^* = true values of GDP, which cannot be observed
m = measured value of nighttime lights
z = a lights-based proxy measure of Y
x = synthetic measure of Y, combining information from both y* and m
i = geographic unit such as country, state, or grid cell
ε_i = measurement error in Y

ξ_i = measurement error in lights
u_i = error in y*-m relationship
α, β, μ = structural parameters

As mentioned above, three basic assumptions are needed: First, there is an unknown true value of GDP for area i, which is measured with error.

$$y_i = y_i^* + \varepsilon_i. \tag{15.1}$$

Second, the lights data is also subject to measurement error due to unobserved physical factors or sampling errors:

$$m_i = m_i^* + \xi_i \tag{15.2}$$

Finally, there is a structural relationship between the observed lights variable and true value of economic output, GDP:

$$m_i = \alpha + \beta y_i^* + u_i \tag{15.3}$$

Equation (15.3) represents a positive relationship between GDP and lights at night. Of course, the relationship could be nonlinear. The exposition below assumes the simplest form of linear relationship.

Because the true values of GDP, y^*, are unknown, the true value of β is unknown. But it can be estimated using observed y and m. The coefficient, $\hat{\beta}$, is a biased estimate due to measurement error in y, that is ε_i in Eq. (15.1). The error-corrected estimate of the structural coefficient, $\widetilde{\beta}$, can be calculated using classical errors-in-variable correction if the variance of measurement error in Eq. (15.1) is known (See Appendix 15.1). Next, the proxy variable can be calculated by inverting Eq. (15.3):

$$\hat{z}_i = \left(1/\widetilde{\beta}\right) m_i, \tag{15.4}$$

where $\hat{z}_i$ is the proxy measure of y^* and $\widetilde{\beta}$ is the corrected coefficient. This proxy measure provides an alternative measure of GDP. The next step in the proposed model explains how proxy variables can be combined with existing measures to improve the precision of estimates of the true values of GDP. It may be best to begin this discussion with an analogy previously given by Chen and Nordhaus (2011): Suppose a person is hiking toward a final destination but does not know its exact location. However, there are separate tools that can be used to help the hiker locate the destination, one being a contour map and another being a GPS device. Although each tool gives the estimate of the exact location of the destination, both estimates are measured with errors which are known to the hiker. If such is the case, the best-guess location of the final destination can be derived if the hiker combines the information from both tools. The error of the best-guess location to the true destination will be lower than the error based on either tool separately. This reasoning can be expressed mathematically.

Assuming the true value of Y, e.g. GDP, is unknown, we can construct a synthetic Y by combing information from observed and proxy measures, that is, taking weighted averages of the two measures of Y:

$$\hat{x}_i = (1 - \theta)y_i + \theta\hat{z}_i. \tag{15.5}$$

Here, $\hat{x}_i$ is a new synthetic measure of Y and θ is the weighting fraction on the proxy measure. The key estimate here is θ. The magnitude of θ indicates the amount of information provided by the proxy variable that can be used to improve the current estimates of GDP. The optimal value of θ can be calculated by minimizing variance of measurement error in the synthetic measure relative to the true values of Y (the proof is presented in Appendix 15.2). Thus, θ is a function of three parameters, σ_ε^2, σ_u^2, and β:

$$\theta* = \frac{\beta^2\sigma_\varepsilon^2}{\beta^2\sigma_\varepsilon^2 + \sigma_u^2} \tag{15.6}$$

Here, σ_u^2 and β can be consistently estimated from Eq. (15.3), and the value of σ_ε^2 can be a prior estimate based on literature. Therefore, the consistent estimator of $\theta*$ can be derived from Eq. (15.6). When the sample size is small, this estimator of $\theta*$ could be biased. Bootstrap techniques can help determine the properties of $\theta*$ estimator for the actual sample (Nordhaus and Chen 2015). In summary, the amount of information that can be derived from a proxy variable such as nighttime lights, is determined by three parameters: (1) the measurement error in y, or observed GDP, (2) the error term and (3) the coefficient in the structural relationship between lights and $y*$ as shown in Eq. (15.3).

With the analytical models of applying proxy variables to improve a conventional measure now established, Sect. 15.3 explains major findings and conclusions from recent studies applying the analytical model to lights measures and the GEcon datasets.

15.3 An Example: Using Nighttime Lights as a Proxy for Economic Statistics

15.3.1 Processing the Two Datasets

Because lights information is stored at a very fine scale, with each pixel corresponding to 30 s of arc, researchers in general aggregate lights pixels according to certain geographic areas suited to their studies. The most common practice is to sum the lights pixel values within administrative boundaries. The following section demonstrates the procedures of aggregating and merging information from lights data and GEcon data. Downloaded from the DMSP archive, nighttime lights imagery files can be processed as raster files in ArcGIS. The GEcon data can be treated as a polygon shapefile. The pixels in light images, in 30 arc-sec, can be summed with the Zonal function by polygons provided either by administrative boundary or grid cell shapefiles. For instance, a complete 1×1 grid cell contains 120×120 pixels. This summation procedure yields around 27,000 terrestrial grid cells for the globe. Time series lights data are available for 17 years, but for some years, two satellites generated separate annual light products, which lead to overlaps of 12 satellite-years. To correct differences across satellites and years, the fixed effects for time and satellites can be used in panel regression estimations. For the analysis at the country level, summed country lights data can be directly merged with country GDP, and for the grid cell analysis, the gridded nighttime lights data can be merged with the GEcon grid cell product by cell ID (available at gecon.yale.edu). For the cells in which country borders cross, the values of nighttime lights and GCP are divided into subcells according to the shares of total cell population in each subcell.

15.3.2 Results of Optimal Weight

Using the analytical model discussed in Sect. 15.1, Chen and Nordhaus (2011) tested how much information nighttime lights can provide as a proxy for national or grid cell economic output. Their study concludes that in general lights data are more useful for poor rather than for wealthy countries and more useful for cross sectional rather than for growth rate measures of GDP and GCP. In the analysis, they assume the relationship between lights and economic output varies by country, and classified countries to a grade system that was initially introduced by Summers and Heston (1991). The A to D country grade is adopted in the current Penn World Table estimates of national output, in which grade A represent developed, high-income countries with a high quality of statistical accounts and with small margins of error in their economic statistics, e.g. Australia, Canada, and the U.S. Grade B, C, and D countries are assigned to countries according to their economic development and qualities of national statistics reported.[2] Grade E, added later, represents the least developed, low-income countries with essentially no statistical systems and with greater measurement error in economic statistics, e.g. Iraq, Mynamar, North Korea (Chen and Nordhaus 2011).

To demonstrate how the light-based proxy variable is generated and its optimal weight is estimated, I use the analysis on the growth rate of grade C countries as an example.[3] There are 103 valid grade-C countries with moderate quality statistical systems. The growth rates of both national GDP and nighttime lights measures over

[2] Some representative countries for grade B are Argentina, Germany, Spain, for grade C are Bangladesh, Egypt, Mexico, Russia, and for grade D are Algeria, Cambodia, D.R. Congo, and Libya.

[3] The detailed list of countries for each grade can be found in the original article by Chen and Nordhaus (2011).

the period 1992–2008 are processed with the steps presented above and merged together.

In this example, a linear regression of the lights growth rate on observed GDP growth rate has a coefficient of .219 ($\hat{\beta}$ in Eq. 15.3), and a mean squared resident of .046 ($\widetilde{\sigma}_u^2$ in Eq. 15.3). Assuming that measurement error of the real GDP growth rate for C countries is 3 percentages per year, one can calculate the error variance of the 17-year GDP growth rate at 0.0153.[4] Since observed GDP is subject to measurement error ε, the coefficient obtained from the above regression, $\hat{\beta}$, is biased, but can be adjusted using σ_ε^2, 0.0153. The formula presented in Appendix 1 is used to calculate the adjusted $\hat{\beta}$, and the corresponding outcome is approximately equal to .25. Next, simply inversing $\widetilde{\beta}$ and multiplying it with the lights growth rate produces the light-based proxy measure of true GDP growth rate, as shown in Eq. (15.4). Whether this proxy can improve the current measures of GDP depends on the magnitude of optimal weight, θ, and three parameters are needed to calculate $\theta - \widetilde{\beta}$, $\widetilde{\sigma}_u^2$, and σ_ε^2 appearing on the right-hand side of Eq. (15.6). The estimates of the first two are directly from the regression of observed lights growth rate on GDP growth rate, and the last parameter, σ_ε^2, is a priori estimates. With this information, the optimal weight on the light-based proxy can be directly calculated as follows:

$$\hat{\theta} = \frac{\widetilde{\beta}^2\sigma_\varepsilon^2}{\widetilde{\beta}^2\sigma_\varepsilon^2 + \widetilde{\sigma}_u^2} = \frac{(0.25)^2 \times 0.015}{(0.25)^2 \times 0.015 + 0.046} \approx 0.02$$

The result shows that the optimal weight is 2 % on light-based proxy and 98 % on measured GDP growth rate. The low weight on the proxy measure suggests that compared to the observed output with moderate accuracy of C countries, the lights data can only provide very limited information to improve the current measure of GDP growth. One of the major conclusions of the above study is that for growth rates, nighttime lights data adds considerable information for D countries, e.g. Algeria, Cambodia, D.R Congo, and Libya, with an optimal weight on the proxy measure around 30 %. However, for A, B, and C countries, where observed national accounts tend to be more accurate, the value added by lights-based proxy is very small, usually less than 3 %. For the cross-sectional economic output, light-based proxy adds a small value for A to D countries, with weights ranging from 1.0 to 12.0 %, but a substantial value for the E countries, with a weight of 25 % for all grid cells (Chen and Nordhaus 2011). The most important finding is that nighttime lights data is found most valuable for the poorest countries or cells with low economic output. Considering the low-quality statistical information and the lack

[4] The calculation of error variance for the 17-year growth rate for grade C countries is 17*(.03^2) = .0153.

of reliable data sources in these areas, the lights data holds huge potential for future research as an alternative data source for studying such areas.

As illustrated above, using weighting estimates is an important improvement in using proxy variables. But, we do not know how reliable this estimates of optimal weight, $\hat{\theta}$, is. Since the value of $\hat{\theta}$ is based on three parameters: the measurement error of observed variables, the coefficient and MSE of the regression equation, the reliability of estimated $\hat{\theta}$ can be determined by the reliability of these three parameters. Using a standard confidence technique in this particular case is questionable due to multiple steps involved in the estimation. An alternative approach, a bootstrap procedure, has been proposed to estimate the precision of weighting parameter, $\widetilde{\theta}_n^*$ (Nordhaus and Chen 2015). Detailed discussion on the bootstrap procedure can be found in Efron and Tibshirani (1986) and Davison et al. (2003). Put simply, a bootstrapping procedure uses data resampling to determine the accuracy of sample estimates. In the nighttime light and economic output studies, Monte Carlo resampling is suggested – the procedure resamples the data with replacement observations and requires the same sample size of the resampled data as was in the original data. Dispersions of the estimates from taking multiple replications can give an overall reliability check on θ* (Nordhaus and Chen 2015).

This empirical application of nighttime lights data in improving estimates of economic output highlights the importance of using formal statistical analysis in utilizing of geocoded information. Almost all previous literature focusing on the relationship between lights and social scientific variables praises the usefulness of lights as a predictor or substitute for conventional variable measures. However, with the analytical models and possible reliability tests proposed above, researchers can easily tell *how much* information can be provided by lights data for improving estimates. With similar approaches, a large swath of remote sensing or geocoded information can be utilized as proxy variables to benefit social scientific investigations in areas where poor quality data is a common problem.

15.4 A Social Demographic Research Application

While mentioned above, it is worth noting here again: One significant implication from the above analysis is that nighttime lights can be used in estimating economic development for poor, underdeveloped regions. Despite early efforts and contributions to expand research into subnational areas of poor countries, Lobao and Hooks (2007) have argued that studies on spatial inequality at the subnational scale have been under-researched. For poor regions, the subnational accounts on inequality or poverty are rare and unreliable, which has limited researchers' investigation and understanding on these topics.

While the first three sections presented above addressed the theoretical and mathematical underpinnings of using geocoded lights data as a proxy variable for economic statistics, this final section will test a theoretical proposition hypothesizing variable relationships in a more traditional social demographic area: the effect of urbanization on poverty rates in developing countries. This section builds off the methodological discussion above and tests whether nighttime lights data can improve the predicting power of urban population for poverty rates of small areas in developing countries. There is both theoretical and empirical evidence on the negative association between urbanization and poverty in developing countries (Ravallion et al. 2007). In the following analysis, first I will use observed subnational urban populations as a traditional measure of urbanization to predict poverty rates. Then, I will create a synthetic measure of urbanization from combining nighttime lights data and conventional urban population measures using optimal weights. The relationship between lights and urbanization is rested on assumptions that people who reside in urban areas are likely to use modern facilities, transportation, and equipment that illuminate the night sky. Thus areas with more urban population tend to be brighter. In the final regression model, initial observed urbanization measures will be replaced by the synthetic measure to test whether a light-based proxy can improve the predictability of urbanization on poverty rates.

The dependent variable of this analysis is small area poverty rates for three poor countries: Vietnam, Nicaragua, and Madagascar. The Center for International Earth Science Information Network for the Poverty Mapping Project (2012) produces the Small Area Estimate (SAE) of poverty rates for thirteen developing countries: Albania, Bangladesh, Bulgaria, Bolivia, Ecuador, Guatemala, Morocco, Madagascar, Malawi, Mozambique, Nicaragua, the West Bank and Gaza, and Vietnam. The SAE poverty rates are consumption-based poverty estimates for subnational areas at either the administrative I or II levels for these countries. In this particular application, I select three countries that have relatively low economic output levels and have SAE poverty rates available at the administrative II level. The particular poverty rate used here is variable *tfgt_0*, defined as the percentage of population who are living under the national poverty line, which is estimated based on national surveys for each country. For Vietnam and Nicaragua, estimates are based on the 1998 national survey, while for Madagascar they are based on its 1993 national survey.

The predictor, urban population rate, is constructed from the urban extent grid and population density grid at 30 arc-sec (1 km) of year 2000 obtained from the Global Rural-Urban Mapping Project at the Center for International Earth Science Information Network (CIESIN 2013). The population data for 30 arc-sec grid cells are aggregated to areas defined by SAE poverty rate maps. The variable urbanization is measured with the ratio of the population residing in urban extent of a subnational area, e.g. county, over total population residing in this area, and is used as the sole predictor in the regression model.

First, I want to test whether the urban population that is measured for the small areas can predict SAE poverty rates for these three countries. The results of this simple regression are reported as Model 1 in Table 15.1. Controlling country dummy variables, urban population can explain about 53 % variance in SAE poverty rate. The regression coefficient is significant at the .001 level, and its

Table 15.1 Dependent variable is SAE poverty rate for Madagascar, Nicaragua and Vietnam

Urban population	Model 1 Observed measure	Model 2 Synthetic measure ($\sigma = .2$)	Model 3 Synthetic measure ($\sigma = .3$)
Regression coefficient	−0.379***	−0.523***	−2.597***
Standardized coefficient	−0.521***	−0.730***	−0.841***
t value	(−22.36)	(−33.12)	(−38.74)
N	868	767	767
adj. R-sq	0.534	0.651	0.713

Country dummy variables are included in all three regression models
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ two-tailed test

standardized form suggests that for every one standardized deviation increase in urban population, the poverty rate declines about a half standardized deviation. It is clear that the urbanization process influences the size of the population living in poverty. We can say that based on information from these three countries the variable urban population provides a moderate amount of information in explaining the SAE poverty rate.

However, the precise measure of the urban population for small areas for these countries is almost nonexistent. Calculating it with urban extent and grid cell population density estimates is undoubtedly subject to measurement error, depending on how the raw data are collected and what methods are used to construct urban extent and grid cell population estimates by CIESIN. So the question is whether we can apply an alternative measure such as nighttime lights data, as a proxy to improve the current urban population measure and therefore obtain more accurate results in the regression analysis?

The analytical model presented in Sect. 15.1 allows me to generate light-based proxy and improved synthetic measure of the urban population variable. Again, according to Eq. 15.6, calculating optimal weights for observed and proxy measures of urban population requires three parameters, the coefficient and MSE from the regression of lights on urban population and a prior estimate of measurement error in the observed urban population variable. First, estimating Eq. (15.3) provides the first two parameters. Controlling country dummy variables, the regression of lights on urban population has a coefficient 4.368 and root MSE of 1.545. Note that the estimated coefficient here is biased due to measurement errors in observed urban population, but it can be adjusted with the formula of classical errors-in-variable correction (Appendix 1). The variance for observed data on urban population is known, .102. If the measurement error of the observed urban population is assumed to be .20, the true variance is .062,[5] and the error-adjusted coefficient is:

[5] Equation for variance of true urban population: $.102 - .20^2 = .062$.

$$\widetilde{\beta} = \left(\frac{\sigma_{y*}^2 + \sigma_\varepsilon^2}{\sigma_{y*}^2}\right)\hat{\beta} = \left(\frac{.062 + .04}{.062}\right)*4.368 \approx 7.181.$$

Hence, according to Eq. (15.6), the optimal weight on the light-based proxy measure is:

$$\hat{\theta} = \frac{\beta^2\sigma_\varepsilon^2}{\widetilde{\beta}^2\sigma_\varepsilon^2 + \widetilde{\sigma}_u^2} = \frac{(7.181)^2 \times 0.04}{(0.7.181)^2 \times 0.04 + (1.545)^2} \approx 0.463.$$

The optimal weight on the observed urban population is .537 (=1 − .463), and a new synthetic measure can be calculated for each record using the above optimal weights:

$$\text{Urban population}_{\text{syn}\_i} = .537 \times \text{urban population}_i + .463 \times \text{lights}_i \times \left(\frac{1}{7.181}\right),$$

where i denotes individual subnational area. Model 2 in Table 15.1 reports the results of the regression using the synthetic measure of urban population as the predictor of SAE poverty rate. Compared to the results of Model 1, the coefficient and its t value and model fit increase significantly in Model 2. The variance in poverty rate explained by the new measure of urban population also increases from 53 to 65 %.

In the above application, the only unknown parameter is measurement error in the observed urban population. When it is assumed to be .20, the obtained optimal weight on light-based proxy is .463. If we assume the error to be .30, the optimal weight on the proxy variable will increase accordingly to .981. It is intuitive that when the observed urban population is subject to a large measurement error, the best guess of true value of urban population relies more on information found in a lights-based proxy variable. Model 3 shows the results of the regression when the error in the observed urban population is assumed to be .30. Compared to the first two models, Model 3 has the higher coefficient of urban population and better model fit, and the variance explained by the new predictor rises to 71 %, about 20 % more than that in Model 1.

This analysis suggests that the synthetic measure built off the light-based proxy can indeed improve the predicting power of urban population on poverty rate. Furthermore, in this particular case, the more information is derived from light-based proxy, the stronger the coefficient and the better the fit the regression model produces. This particular application suggests that lights can provide good supplemental information for measuring urban population of small areas for these three countries. The fact that the improvement of the regression model is closely associated with the measurement error in observed data is especially important for future social demographic studies, since for many poor countries where formal statistical data are sparse, estimated social demographic variables, especially estimates at the subnational level, tend to be subject to large measurement errors. Thus, using satellite-based nighttime lights data as a proxy can directly solve these existing problems of poor quality in data estimated with traditional survey methods or derived from national accounts and provides better estimates for analysis focusing on underdeveloped world regions or smaller areas.

15.5 Conclusions

This chapter aimed to introduce important statistical models and important geocoded datasets measured with unconventional methods that could be applicable for mid-range studies. It also incorporated empirical applications based off these models and data which could benefit researchers of demography and other social science disciplines in various ways. First, there is a data shortage for a wide range of socio-economic and demographic variables in less developed regions. Geocoded data, including remote sensing information, provides potential solutions to this problem. An excellent example is satellite-based nighttime lights data. Such data is available for the entire globe and is available at a very fine scale, which makes it valuable in investigating not only cross-country differences, but also internal variation within countries and spatial dependency at different scales. The choice of scale or aggregation and spatial dependency has been viewed as important issues in spatial demography (Parker 2014). With rich information provided by lights data, those topics can be further explored.

Second, the empirical applications presented in this chapter set examples for future studies that could benefit from using nighttime lights data as proxy variables. While the first application demonstrates the feasibility of using the model and lights data in estimating economic development, the second application shows that improved measures based on the lights proxy can be used directly in hypothesis testing processes. There is little doubt that statistical modeling with nighttime lights as proxy measures will expand future research. However, such proxy data, as forwarded here, probably benefits most that research focusing on subnational or small area phenomena in low-income regions where lack of or poor quality data has severely hindered social scientific investigations up to this point.

Finally, although many studies have explored the relationship between lights and other variables, very few have adopted formal statistical models that test the precise amount of information that can be derived from lights data. The analytical models introduced in this chapter not only provide mathematical calculations of optimal weights on proxy measures, but also opens it to formal reliability testing and sensitivity analysis. The methodology presented here allows lights to be tested as proxy for many other social demographic variables, but more importantly, it can be applied to test many other types of geocoded data, including a wide range of remote sensing information that has not been fully utilized by social scientists.

Fifteen years ago, Rindfuss and Stern (1998) discussed the need and challenges of linking remote sensing and social science research. Their insightful comments along with other early applications of remote sensing published in *People and Pixels* (Moran et al. 1998) represent seminal work in this field. With unprecedented increases in geospatial information collected from governmental, private and academic institutions, we are entering a new era that can potentially revolutionize social science research if we can use such information efficiently and prudently. This chapter has attempted to provide some theoretical foundations and initial applications in how such information may be utilized in social scientific research.

Appendices

Appendix 15.1

The formula for corrected coefficient using classical errors-in-variable correction:

$$\widetilde{\beta} = \left(\frac{\sigma_{y*}^2 + \sigma_\varepsilon^2}{\sigma_{y*}^2}\right)\widetilde{\beta}$$

$\widetilde{\beta}$ is the estimated coefficient from the regression model; σ_ε^2 is the a priori estimate of error variance of true value of Y, $y*$, and σ_{y*}^2 is variance of $y*$.

Appendix 15.2

The mean squared error (MSE) of $\hat{x}_i$, $V(\theta)$, is a function of the weight, θ:

$$\begin{aligned} V(\theta) &= E[(1-\theta)y + \theta\hat{z} - y*]^2 \\ &= E\left[(1-\theta)(y* + \varepsilon) + \tfrac{\theta}{\beta}(\beta y* + u) - y*\right]^2 \\ &= E\left[(1-\theta)\varepsilon + \tfrac{\theta}{\beta}u\right]^2 \\ &= (1-\theta)^2\sigma_\varepsilon^2 + \frac{\theta^2}{\beta^2}\sigma_u^2 \end{aligned}$$

Minimizing $V(\theta)$ with respect to θ yields the optimal weight, $\theta*$, as a function of three parameters, σ_ε^2, σ_u^2, and β:

$$V'(\theta*) = 0 = -2(1-\theta*)\sigma_\varepsilon^2 + \frac{2\theta*}{\beta^2}\sigma_u^2$$

Or Eq. (15.6):

$$\theta* = \frac{\beta^2\sigma_\varepsilon^2}{\beta^2\sigma_\varepsilon^2 + \sigma_u^2} \tag{15.6}$$

References

Balk, D., Montgomery, M. R., McGranahan, G., Kim, D., Mara, V., et al. (2009). Mapping urban settlements and the risks of climate change in Africa, Asia and South America. In J. M. Guzman, G. Martine, G. McGranahan, D. Schensul, & C. Tacoli (Eds.), *Population dynamics and climate change* (pp. 80–102). New York: IIED/UNFPA.

Baugh, K., Elvidge, C., Ghosh, T., & Ziskin, D. (2010). Development of a 2009 stable lights product using DMSPOLS Data. In *Proceedings of the 30th Asia-Pacific Advanced Network meeting* (Vol. 30, pp. 114–130).

Buhaug, H., Gleditsch, K., Holtermann, H., Østby, G., & Tollefsen, F. (2011). It's the local economy, stupid! Geographic wealth dispersion and conflict outbreak location. *Journal of Conflict Resolution, 55*(5), 814–840.

Cederman, L. E., Weidmann, N., & Gleditsch, K. (2011). Horizontal inequalities and ethno-nationalist civil war: A global comparison. *American Political Science Review, 105*(3), 478–495.

Center for International Earth Science Information Network – CIESIN – Columbia University. 2013. *Environmental Treaties and Resource Indicators (ENTRI) Query Service*. Palisades: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/entri. Accessed 19 June 2013.

Center for International Earth Science Information Network (CIESIN). (2013). *Global rural-urban mapping project, version 1 (GRUMPv1): Urban extents grid and population density grid*. http://plue.sedac.ciesin.columbia.edu/plue/ddviewer/. Accessed 3 Jan 2010.

Chen, X., & Nordhaus, W. (2011). Using luminosity data as a proxy for economic statistics. *The Proceedings of National Academy of Sciences, 108*(21), 8589–8594.

Davison, A. C., Hinkley, D. V., & Young, G. A. (2003). Recent developments in bootstrap methodology. *Statistical Science, 18*(2), 141–157.

DeFries, R., Pagiola, S., Adamowicz, W. L., Akcakaya, H. R., Arcenas, A., et al. (2005). Analytical approaches for assessing ecosystem condition and human well-being. In R. Hassan, R. Scholes, & N. Ash (Eds.), *Ecosystems and human well-being: Current state and trends* (Vol. 1, pp. 37–71). Washington, DC: Island Press.

DMSP-OLS Nighttime Lights. (2013). *Image and data processing by NOAA's National Geophysical Data Center*. DMSP data collected by the US Air Force Weather Agency. http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html. Accessed 10 Jan 2011.

Doll, C., Muller, J., & Elvidge, C. (2000). Nighttime imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *Ambio, 29*, 157–162.

Ebener, S., Murray, C., Tandon, A., & Elvidge, C. (2005). From wealth to health: Modeling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal Health Geographics, 4*, 5–14.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1*(1), 54–77.

Elvidge, C., Baugh, K., Kihn, E., Kroehl, H., Davis, E., et al. (1997). Relation between satellites observed visible – Near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing, 18*, 1373–1379.

Elvidge, C., Imhoff, M., Baugh, K., Hobson, V., Nelson, I., et al. (2001). Night-time lights of the world: 1994–1995. *ISPRS Journal of Photogrammetry and Remote Sensing, 56*, 81–99.

Elvidge, C. D., Baugh, K. E., Kihn, E., Kroehl, H., Davis, E. R., & Davis, C. W. (1997). Relation between satellites observed visible – Near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing, 18*, 1373–1379.

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Hobson, V. R., Nelson, I., Safran, J., Dietz, J. B., & Tuttle, B. T. (2001). Night-time lights of the world: 1994–1995. *ISPRS Journal of Photogrammetry and Remote Sensing, 56*, 81–99.

Elvidge, C., Safran, J., Tuttle, B., Sutton, P., Cinzano, P., et al. (2007). Potential for global mapping of development via a nightsat mission. *GeoJournal, 69*, 45–53.

Elvidge, C., Sutton, P., Ghosh, T., Tuttle, B., Baugh, K., et al. (2009). A global poverty map derived from satellite data. *Computers & Geosciences, 35*(8), 1652–1660.

G-Econ. (2013). *Geographically based economic data*. Department of Economics, Yale University. http://gecon.yale.edu/. Accessed 19 Oct 2011.

Henderson, V., Storeygard, A., & Weil, D. (2012). Measuring economic growth from outer space. *American Economic Review, 102*(2), 994–1028.

Lobao, L., & Hooks, G. (2007). Advancing the sociology of spatial inequality: Spaces, places, and the subnational scale. In L. Lobao, G. Hooks, & A. Tickamyer (Eds.), *The sociology of spatial inequality* (pp. 29–63). Albany: State University of New York Press.

Moran, E. F., Rindfuss, R. R., & Stern, P. C. (1998). *People and pixels: Linking remote sensing and social science* (pp. 56–59). Washington, DC: National Academy Press.

Noor, A., Alegana, V., Gething, P., Tatem, A., & Snow, R. (2008). Using remotely sensed nighttime light as a proxy for poverty in Africa. *Population Health Metric, 6*(1), 5.

Nordhaus, W. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3510–3517.

Nordhaus, W., & Chen, X. (2015). A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. *Journal of Economic Geography, 15*(1), 217–246.

Nordhaus, W., Azam, Q., Corderi, D., Hood, K., & Makarova, N., et al. (2006). *The G-Econ database on gridded output: Methods and data*. Yale University. http://gecon.yale.edu/sites/default/files/gecon_data_20051206.pdf. Accessed 3 Nov 2011.

North, G. R., Biondi, F., Bloomfield, P., Christy, J. R., Cuffey, K. M., et al. (2006). *Surface temperature reconstructions for the last 2,000 years*. NRC Statement to Subcommittee on Oversight and Investigations, Committee on Energy and Commerce, US House of Representatives. http://www7.nationalacademies.org/ocga/testimony/surface_temperature_reconstructions.asp. Accessed 22 Nov 2011.

Parker, D. M. (2014). Human migration and spatial synchrony. In F. Howell, J. Porter, & S. Matthews (Eds.), *Recapturing space*. New York City: Springer.

Ravallion, M., Chen, S., & Sangraula, P. (2007). New evidence on the urbanization of global poverty. *Population and Development Review, 33*, 667–701.

Rindfuss, R. R., & Stern, P. C. (1998). Linking remote sensing and social science: The need and the challenges. In E. F. Moran, R. R. Rindfuss, & P. C. Stern (Eds.), *People and pixels: Linking remote sensing and social science* (pp. 1–27). Washington, DC: National Academy Press.

Sanderson, E. W., Jaiteh, M., Levy, M., Redford, K., Wannebo, A., et al. (2000). The human footprint and the last of the wild. *Bioscience, 52*, 891–904.

Seo, S. N. (2011). The impacts of climate change on Australia and New Zealand: A gross cell product analysis by land cover. *Australian Journal of Agricultural and Resource Economics, 55*, 220–238.

Summers, R., & Heston, A. (1991). The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988. *The Quarterly Journal of Economics, 106*, 327–368.

Sutton, P., Elvidge, C., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal Ecological Economics and Statistics, 8*, 5–21.

Tang, M., & Woods, D. (2008). The exogenous effect of geography on economic development: The case of sub-Saharan Africa. *African and Asian Studies, 7*, 173–189.

Tatem, A., Adamo, S., Bharti, N., Burgert, C., Castro, M., Dorelien, A., et al. (2012). Mapping populations at risk: Improving spatial demographic data for infectious disease modeling and metric derivation. *Population Health Metrics, 10*(8), 14.

The Poverty Mapping Project. (2012). Center for International Earth Science Information Network (CIESIN), Columbia University. *Global subnational infant mortality rates and SAE poverty rate dataset*. CIESIN, Palisades, NY, USA. http://sedac.ciesin.columbia.edu/. Accessed 15 Mar 2012.

Ziskin, D., Baugh, K., Hsu, F.-C., & Elvidge, C. D. (2006). Methods used for the 2006 radiance lights. *Proceedings of the Asia-Pacific Advanced Network, 30*(2010), 131–142.

Chapter 16
Human Migration and Spatial Synchrony: Spatial Patterns in Temporal Trends

Daniel M. Parker

16.1 Introduction

Demographic research is almost always focused on temporal processes in populations. At its core, demography is concerned with changes in populations, frequently but not always focusing on fertility, mortality, and migration. These changes take place over some unit of time, meaning temporal dynamics are inherent and quite often are explicitly modeled in demographic analyses. However, these processes don't take place in the absence of geography and demographers aren't only interested in temporal dynamics, they are increasingly interested in spatial dynamics too (Matthews and Parker 2013; Porter and Howell 2012; Voss 2007; Wachter 2005). Things that are close in geographic proximity are often more alike than things which are geographically distal, and temporal trends are typically tied to the spaces and places in which they occur (Tobler 1970). Population densities in geographically proximate populations may rise and fall in synchronous fashion or conversely may have opposing patterns altogether. Mortality rates in neighboring regions can be extremely similar as can fertility and migration rates. This spatial synchrony or spatial covariation may be the result of several factors, including the socio-cultural characteristics of a region, the natural or built environment, or even epidemiological factors.

Perhaps migration offers one of the most readily available examples of demography as an inherently spatial science. Migration is unique as a topic of interest to demographers as it is necessarily a spatial process (Wachter 2005). That is, in order to migrate, one must physically move from one place to another. This spatial process is also somewhat defined by its temporal component. Some migrants

D.M. Parker, Ph.D (✉)
The Population Research Institute, Center for Infectious Disease Dynamics, The Pennsylvania State University, University Park, PA 16802, USA
e-mail: Daniel.P@tropmedres.ac

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_16

move permanently while on the other end of the spectrum, many people make almost daily physical moves to another location (e.g. to school, to work, etc). Both long and short term migrations[1] occur heterogeneously across populations, with different age or sex groups moving at different rates, distances or even directions, and for different reasons. Furthermore, migration *rates* may cluster within or across regions, populations, and subpopulations (e.g. ethnicity or age groups). For a host of reasons, migration is therefore intrinsically wedded to spatial demography.

Merging theory and empirical data is important for any scientific discipline. Theory in demography differs in its approach when compared to theory in other disciplines, perhaps because it describes general processes rather than attempting to explicitly describe why those processes occur at an individual basis. Some have even claimed that demography is a method without a theory (see Burch 2003). Perhaps, however, this comes from a narrow view of what theory is, as demography explains, usually through models, the ways that populations behave and the age schedules that are associated with life events in populations (Burch 2003). Malthus' theory of population growth, Lotka's model of stable populations, Thompson's demographic transition model, and Henry and Coale's work on fertility in "natural" populations are all examples of demographic theory (Malthus 1817; Lotka 1998; Thompson 1929; Coale 1971; Henry 1961). Demography does have theory; however that theory differs from the theory that most demographers learn in their sister, social science disciplines which are typically focused on the components of populations rather than the whole.

That most students of demography are dually trained, both in demography and in another social science discipline, also makes demographers and demography unique because the field is inherently multidisciplinary. Demographers also have much in common with population ecologists and population biologists. And while populations are a topic of interest that shouldn't have to be married to another discipline in order to warrant study,[2] this common arrangement leads to a deep understanding of the processes being studied. Demography helps us understand population processes whereas our social science backgrounds help us understand, from more micro-level standpoint, *why* people do the things that they do. Ultimately, demographers are well-suited for deriving explanations for overarching population processes, and perhaps especially for merging theory with empirical data.

I therefore begin this chapter by drawing on some theory behind the mechanisms and causes of spatio-temporal population dynamics, using migration as a substantive topic. I look at both age-specific and repeated cross sectional data as two different focal points for understanding human migration and movement patterns.

[1] Short-term migrations may also be referred to as human mobility or circular migration and in population ecology parlance, dispersal.

[2] Nathan Keyfitz, arguably among the most influential demographers, was a strong proponent of this line of thought. While his PhD was in Sociology, he collaborated with population scientists from quite a wide array of disciplines.

While short term movements aren't typically discussed at the same time as migration, I am here considering all ranges of human movement, or dispersal in population ecology parlance, with permanent migration being a special subcategory of human movement. In order to illustrate the merging of empirical data and theory, I then draw on data on a highly mobile ethnic group (the Karen) in Southeast Asia. Then I close with a discussion on some new and old issues related to spatial demography, namely: issues of scale, new forms of data and how to deal with them, and of course, ethical considerations.

16.1.1 Spatial Synchrony

Populations that are related geographically can simultaneously be affected by macro-scale processes. For example, a contraceptive policy rolled out in a developing region might lead to decreasing fertility throughout that region. Separate, yet geographically proximate subpopulations within that overarching region may exhibit strong synchrony in decreasing fertility rates even if other factors vary in those subpopulations. Furthermore, we might expect to see decreasing synchrony between subpopulations that are further and further away from each other, especially since at some point we would be comparing populations that are no longer covered under the same contraceptive policy. Essentially, we might expect to see time-series trends that are very similar for populations that are more proximate than those which are more geographically separated. Similar examples could be given using meteorological factors (rain and ambient temperature) in agricultural populations, epidemiological landscapes (which are also influenced by environmental factors), macro-scaled economic policies, political factors, and educational schedules, all of which can and likely do influence human demography.

Given a long set of observations (time series data) there are a variety of methods, both parametric and nonparametric, for measuring correlations between distance and synchrony (Liebhold et al. 2004). Early methods for assessing synchrony included plotting and visually inspecting data. Another simple method is to look for correlations between time series data, using Pearson's product-moment or Spearman's correlation coefficients, or the lag-0 cross-correlation coefficient. Linear decay models have been used to investigate decreasing synchrony with increasing distance. A nonparametric model might be more suitable to complex distance – synchrony relationships. For example, Bjørnstad & Bolker created a method which compares smoothing splines across landscapes (Bjørnstad and Bolker 2000). Methods for detecting and measuring more complicated forms of spatial synchrony have also been created. For example, a plot of synchrony over distance sometimes reveals waves in distance-synchrony relationships (Lieberman 1993). Such waves can also move over time, leading to a phenomenon referred to as travelling waves, of which the speed and sometimes points of origin can be determined. Such traveling waves have been applied to several substantive areas of research, including the epidemiology of measles, dengue, and influenza, as well as other host parasite and predator prey relationships (Stark et al. 2012; Cummings et al. 2004; Grenfell et al. 2001; Bjørnstad et al. 1999). Some researchers have also used phase analyses to calculate correlations and measure lags between peaks in time series data (Cazelles and Stone 2003).

There are both *mechanisms* and *causal* factors that can lead to such synchrony. With regard to mechanisms, Moran[3] showed early on that two populations with the same density dependent relationships (i.e. the relationship between population growth and population density is the same) can exhibit population synchrony through correlated density independent factors (such as meteorological events) (Hudson and Cattadori 1999; Bjørnstad et al. 1999; Moran 1953; Royama 1992). The so called "Moran effect" suggests that the correlation between the population densities (p_d) of two populations is equal to the correlation in their environments (p_e): $p_d = p_e$. Moran was concerned with the population dynamics of the Canadian lynx (*Lynx canadensis*) with regard to meteorological factors, however it is not hard to imagine situations in which density independent factors (several of which I previously mentioned) influence the dynamics of human populations as well. Conversely, dispersal may also influence populations by making them more homogeneous, though this factor can be complicated by assimilation and acculturation in humans (Ranta et al. 1995). Some studies have shown that while migrants often arrive in a new nation with the fertility rates of their nation of origin, they quickly assimilate to the fertility rates of their destination (Parrado and Morgan 2008) (but also see (Frank and Heuveline 2005)). The same has also been shown with regard to health (Abraído-Lanza et al. 1999).

The causes of such synchrony are multifold and may be extremely complex. However, while causal factors in the synchrony of population dynamics have been among the most vexing of issues for population ecologists, some of the factors leading to spatial covariance and population synchrony in human populations are intuitive. Socio-cultural, economic, and political factors that influence population dynamics can be inferred through social science studies. The trick then becomes wading through the overabundance of information, the numerous different potential contributing factors, to arrive at an understanding of what led to the population dynamics of interest. However, when such dynamics co-vary spatially, we should at least be able to narrow down those factors which are shared by populations with shared dynamics. Are there external density independent factors that are shared by populations? Or are the populations in fact virtually the same because of the mixing of ideas and even population members?

Models that incorporate both space and time in demography have historically been lacking. Most explicitly incorporate time or space at the expense of the other. Arguably, models that incorporate a spatial component are far outnumbered by those with temporal components in demographic research. Spatial synchrony can

[3] Patrick Alfred Pierce Moran also developed the Moran's I statistic, a commonly used measure of spatial autocorrelation.

perhaps offer a theoretical and methodological bridge, by simultaneously linking space and time, when spatially referenced time series data are available.

16.1.2 Is Migration Law-Like? Mathematical Descriptions of Migration

Demography is now mostly a statistical science, but at least some of the founders of modern demography saw it as a mathematical science too[4] (Burch 2011; Romaniuk 2011). Consider Lotka's mathematical theory of stable populations and Keyfitz's contributions with regard to sensitivity analysis, population dynamics, and applied demography (Keyfitz and Caswell 2005; Lotka 1998). Another place where demographers have incorporated mathematical methods is in modeling age patterns of demographic events such as mortality, fertility, and finally migration; applying mathematical functions that describe the shape of age-specific curves in demographic processes.

Some life events have such strong age-specific characteristics that it makes sense to create a model schedule of events for comparisons and investigations into deviations from the norm. Mortality is probably the best studied of these model schedules, however Rogers and Castro (1981) extended this practice to migration (Rogers and Castro 1981). Generalizing from the age profiles of internal migrants from several Western world settings, they created a set of four multiexponential model migration schedules. In the general model there are 11 parameters, 7 of which specify the shape of the migration curve by age and 4 that describe the intensity of migration by age. The most common components of these models (see Fig. 16.1) describe the decrease in high migration from early childhood (when children are moving with their parents in the formation of new households) and a rise in migration during the working and marrying ages. Some models also have "retirement" peaks in later age groups (Raymer and Rogers 2008).

Age-specific rates of migration aren't the only thing that varies. The different components of these model schedules can be considered individually with regard to whether or not the reasons for migration in each component are the same as well as whether or not they consist of dependent (as with migration in very young ages) or independent (occupational migration) migrants. An analysis of the life course can therefore inform or explain a given schedule of migration. Finally, while this general pattern in age-specific migration may approach a socio-cultural universal, there are likely to be subtle differences in the migration schedules of different regions of the world, which have different population structures, different societal norms, and different economic and political factors.

[4] In fact, Lotka saw the study of human populations as being divided into mathematical demography and statistical demography.

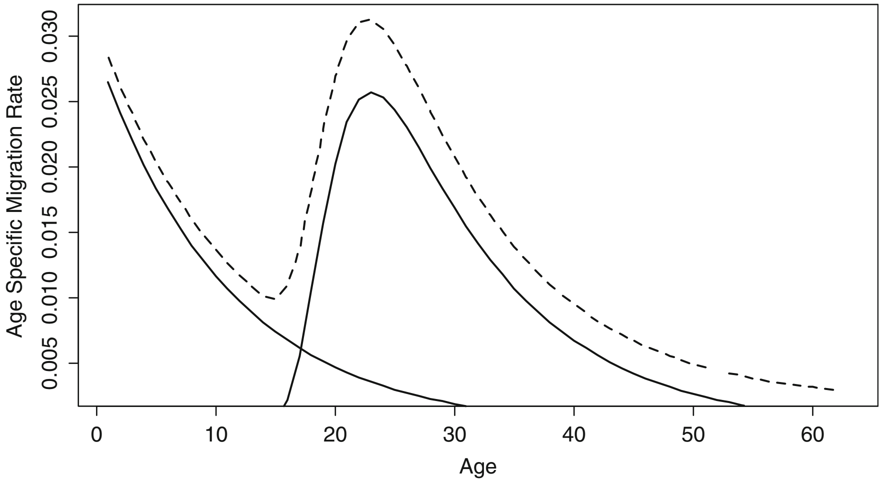

Fig. 16.1 A typical model schedule for age-specific migration from Stockholm 1974 (Reproduced from Rogers and Castro 1981)

16.1.3 Temporal Patterns in Human Migration and Movement

Humans (and other animals) make permanent, seasonal, monthly, even daily or hourly movements. Sometimes there are rhythms in those movements, with schedules based on seasons, daylight, and holidays leading to synchronous movements within and across populations. With regard to short-term movements, consider rush-hour traffic in which many people leave their households to spend much of the day in a location that may be many miles away from their household. Later in the afternoon swarms of people leave work to return home. College students may leave home in early fall and return in the winter. Agricultural workers follow seasonal harvests for their livelihoods, repeating harvesting cycles year after year.

Each of these suggested movement patterns are driven by macro-scale factors such as daylight and seasons, which are shared across large geographical spaces. This means that movement patterns may appear synchronous across those geographical spaces which share the same sunrise and sunset times, the same seasonal patterns, the same time zones, or across regions with shared holidays[5] and a number of other factors. While there are always exceptions to these trends and synchronies, these general patterns are quite predictable over time.

[5] Some holidays are highly localized (consider Juneteenth in Texas or Patriots day in Massachusetts and Maine), whereas others have an extremely broad range (i.e. New Year's day, Christmas, etc.).

Space and time also intersect in other interesting ways, for example, with regard to interactions between frequency, distance, and travel times[6] (Hägerstrand 1970). Trains, airplanes, highways, and automobiles have made it much easier to travel farther distances in shorter amounts of time. But constraints do remain. Travel distance is directly related to travel time, though the relationship between these two factors has changed much over the last couple of centuries. Travel frequencies aren't likely to be very high when travel time is very long (Hägerstrand 1970). For most people, short term movement patterns such as those that occur daily, are likely to occur relatively near the home (Wang et al. 2011). More long term patterns aren't quite as constrained by this trend, and permanent moves are likely at this point to be more constrained by social networks, information, and economic circumstances than physical ability (Pred 1977). While the ability to literally travel around the world is now much easier than it once was, doing so frequently remains unlikely.

16.1.4 Why Does Migration Happen?

Migration occurs for a lot of very complex reasons and a comprehensive examination of the motivations, drivers, and reasons behind migration is beyond the scope of this chapter. However, as previously mentioned, demographic analyses are frequently supported by drawing on social science theory in order to attempt to explain *why* people do what they do. In that spirit, I briefly touch on several theoretical explanations for migration motivations, but will only be able to scratch the surface. Early models for explaining migration focused on large areal units and comparative inequalities that either pushed would-be migrants away from regions with scarce materials or pulled them to regions that are relatively rich in materials (Fields 1976; Muth 1971). However, researchers noted the extreme heterogeneity in migration processes, even within the same migrant groups, destinations, and places of origin. Not all people migrate, not all migrants follow what appear to be logical migrant streams. The focus therefore shifted to individual-level theories that were still primarily focused on economic factors (Stark and Bloom 1985; Todaro 1980). Most of these models looked at the decision to migrate in a cost-benefit balancing type equation, where the benefits (real or perceived) outweigh the costs of making a move. However, while individual migrants can and do act as their own agents, they are also embedded in households, families, neighborhoods, communities, villages, and regions (De Jong and Gardner 1981; Portes and Sensenbrenner 1993). The context in which an individual would-be migrant lives, or is exposed to, matters in migration decisions. Context can be cumulative and can change over the course of a lifespan, as can individual wants and needs, meaning that a life course approach to understanding migration decisions is valuable (Howell and Frese 1983; Howell

[6] There is a literature concerning "Time Geography" that is relevant to this and is of general interest with regard to space-time interactions.

1981). Further extensions in migration studies have considered social and cultural factors, as well as gender, in migration processes (Curran and Saguy 2001; De Jong et al. 1996). Each of these aspects may influence not only the decision to migrate, but also who migrates, where they go, when or if they will return, if they send home remittances, etc. Similar factors also influence general human movement patterns at very short (daily) or seasonal patterns (Winterhalder and Smith 1992; Wang et al. 2011).

Many of these considerations are arguably more about place than space, though there are clear interactions between the two. That is, the characteristics of a place can influence the migration situation in that place. As alluded to above, areas with relatively poor natural resources might act as “push” factors, leading people who live in such areas to migrate out at higher rates than in other regions. People living in unsavory conditions, for example, in areas with high crime, with warfare, with poor health conditions, etc. may also be more likely to at least attempt to leave those places. However, social scientists will note that this effect can be extremely heterogeneous, with some individuals remaining in such areas despite it seeming logical to leave.

Another related consideration that is perhaps more commonly discussed among mathematical demographers and population ecologists concerns density dependence in migration. More than a few studies have looked at the potential for population density to influence all sorts of population processes (Caswell 2006; Relethford 1986; Wood et al. 1985). For example, regions with high population densities may have low per capita resources, leading to higher rates of out-migration, sometimes referred to as Allee effects (Allee and Bowen 1932). Some researchers have even proposed that finding marital partners will be more difficult under some density situations, which could also lead to either in- or out-migration (Swedlund 2009; Mielke et al. 1994). While such factors can influence human migration, they aren’t likely to be completely deterministic. Sociocultural factors can be and are extremely powerful.

For example, post-marital residence rules influence out-migration and dispersal patterns. Historically speaking, most people did not move very far from their place of origin. Marriage was typically with someone from nearby and neolocal households didn’t stray very far from the couples’ houses of origin (Coleman and Haskey 1986; Harrison 1995; Fix 1999; Wijsman and Cavalli-Sforza 1984). Transportation has changed this pattern, to some extent and especially in some societies, quite drastically (Harrison 1995). Virilocal[7] post-marital residence patterns, in which newly married couples live either with or very near the male spouse’s family, will lead to interesting patterns in population dynamics, perhaps especially spatially. Most populations practice some form of exogamy, meaning that there are rules about who to marry and marriage with people too close to an inner circle is taboo.

[7] Here I use the terms virilocal and uxorilocal rather than patrilocal and matrilocal, respectively. My reasoning is that the terms patrilocal and matrilocal assume a unilineal descent system, which is not always the case.

At the same time, most populations also practice some form of endogamy. Most people don't marry people that are too different from themselves. Therefore we might expect to find situations in which individuals choose to marry people that don't live too far away from their home or family of orientation, but far enough so as not to be breaking incest taboos. In a virilocal society, then, females who out-migrate for marital reasons are likely to move further away than are males. The opposite case would occur in uxorilocal societies, with dispersal being more geographically widespread for males than for females.

16.2 Spatial and Temporal Variation in Karen Migration

16.2.1 Ethnic Karen Along the Thai-Myanmar Border

The Karen are the largest ethnic minority in Thailand, and one of the largest ethnic groups in Myanmar (formerly Burma) (Rajah 2008). They are primarily subsistence agriculturalists, meaning that they are largely dependent on the rice that they grow, though they are far from isolated from the Thai economy. Their dependence on agriculture means that work is highly seasonal, with high intensity working periods at the beginning and end of the rainy season, corresponding to land clearing, planting, and harvesting crops(Kunstadter 1972, 1983).

The Karen have also been engaged in low-intensity warfare with the Myanmar government (Lee et al. 2006) for around half a century but have been at peace for the last year Many Karen fled their ancestral lands in Myanmar for the relative safety of Thailand. In some cases, thousands of Karen refugees have flooded across the border in short periods of time, crowding into large refugee camps along the Thai-Myanmar border (TBC 2004). There are also much smaller movements, with individuals, families, and villages relocating across the border.

In all cases, there are reasons to return to Myanmar for various amounts of time. For example, many Karen have family and friends remaining on the Myanmar side of the border. Holidays and down times in the agricultural year are times when Karen from Thailand may travel back to Myanmar to celebrate and to visit friends and family. Furthermore, some Karen actually farm plots of land in Myanmar. The international border is only marked by a relatively small river in many regions, and unused land on the Myanmar side of the border can offer a source of rich soils for farming.

Finally, some movements among adolescents and children are related to schooling. As agriculturalists, the Karen of Thailand mostly inhabit rural areas in Northwestern Thailand. Schooling is important to the Karen, however there are few local schools past age 14, when secondary schooling begins. This means that many children travel to regional secondary schools and frequently stay in school dorms for the duration of the semester. As previously mentioned, migration motivations and dynamics may vary by age group. In the case of adolescents in secondary school, there is seasonal migration that does not exactly correspond to seasonal migration in adults who are tied to the land.

The long international border (over 2000 km) between Thailand and Myanmar is difficult to police, and the region is known as a hub for the illegal trade of narcotics (the "Golden Triangle"), precious gems, lumber, cattle, and humans (Lintner 2000). This means that there is a relatively large amount of clandestine movement across the border, not only by merchants in the black market economy but also by Karen who simply walk across the unguarded border to visit Myanmar. It also means that for many reasons this place, which has no well-defined borders, provides a contextual environment that is highly conducive to heavy flows of human migrant traffic. Much of this movement is temporary, but there can be a lot of variation in the duration of travels. Since the Karen are tied to the land because of their dependence on agriculture, much of this movement can be expected to occur seasonally.

16.2.2 Analysis and Results of Karen Out-Migration

My research among the Karen of Thailand looks at the influence of household and individual-level factors on demography and epidemiology. In 2011, as part of a large, National Institutes of Health initiative for understanding and controlling malaria in this region, demographic and epidemiological surveys were undertaken in each of four Karen villages along the Thai-Myanmar border. The villages are arranged from North to South with about 35 km separating the northernmost from the southernmost village (Fig. 16.2, Table 16.1). Full censuses are taken twice

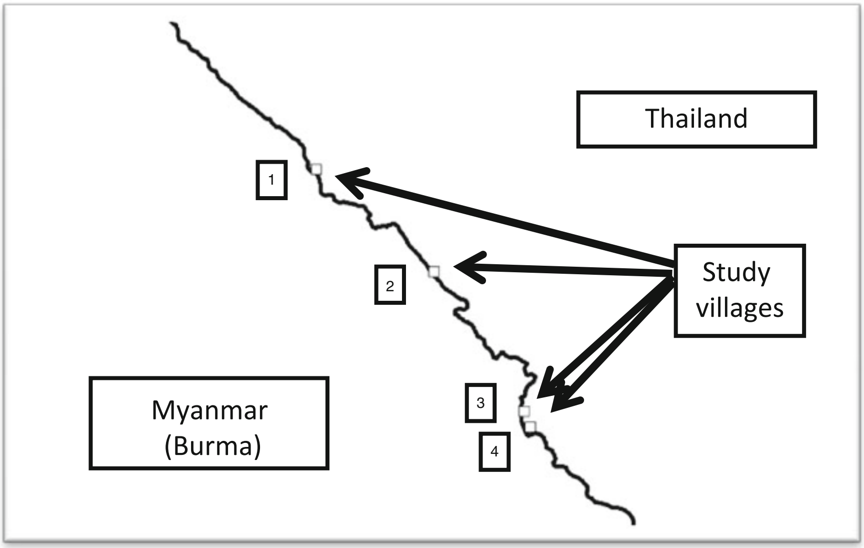

Fig. 16.2 Study villages

Table 16.1 Village characteristics

Village	Altitude	Population	Households	People/House	% Female
1	96	512	108	4.74	52
2	111	798	145	5.50	48
3	132	1194	232	5.15	49
4	128	2073	373	5.56	51

yearly and every other week a mobile health team moves through the villages, going house-to-house, asking about any new additions to households (in-migrations or births) as well as any missing individuals from households (out-migrations and deaths). Those who have moved out of a household and aren't expected to return within a month are coded as out-migrants. These are the individuals (out-migrants) who are in the subject of the following analysis, with time-series data for 13 months.

The temporal component of migration is multi-fold: migration may occur more frequently during certain times during the life-span as well as during certain times of the year. We can therefore think about temporal trends in migration from both cohort (age-specific) and repeated cross-sectional perspectives. While there are general trends in both sets of temporal dynamics (age and the calendar year), the difference between the two is striking. Age schedules of migration strongly conform to each other, whereas the time series trends are more loosely coupled. That is, correlations are much stronger between the life span data of the four study villages when compared to the time series migration data from those same villages.

From Fig. 16.3 we can see a small dip in out-migrations in children approximately 5 years old, followed by a steep rise in migration rates among those around age twenty. Afterward, migration rates slowly taper off throughout the rest of the life span. Comparisons across the villages indicate that peak migration rates are lower in the two villages with smaller overall population sizes, but that the temporal dynamics are largely mirrored across all four villages. This could potentially indicate some density dependence in out-migrations with regard to age. At least some previous research has indicated density dependence in migration rates (Relethford 1986; Umezaki and Ohtsuka 2002; MacDonald and Hewlett 1999; Mielke et al. 1994; Wood et al. 1985). Table 16.2 lists pairwise Spearman's rank correlation coefficients and Pearson's product-moment correlation coefficients for age-specific rates of out-migration between the villages, indicating high correlations between all combinations of the four villages. *P*-values for these correlations are likely to be unreliable because of both spatial and temporal dependence.

Conversely, the seasonal dynamics of out-migration appear to vary more widely across villages, perhaps especially among those which are furthest apart geographically (Fig. 16.4). The two villages which are closest together (3 and 4) appear to have the most consistently correlated trends (from Fig. 16.4 and Table 16.3), followed by village 2 and 1. I then created a distance matrix based on spatial coordinates (village centroids), a difference matrix based on average village elevation, and correlation matrices for both age-specific and time series out-migration

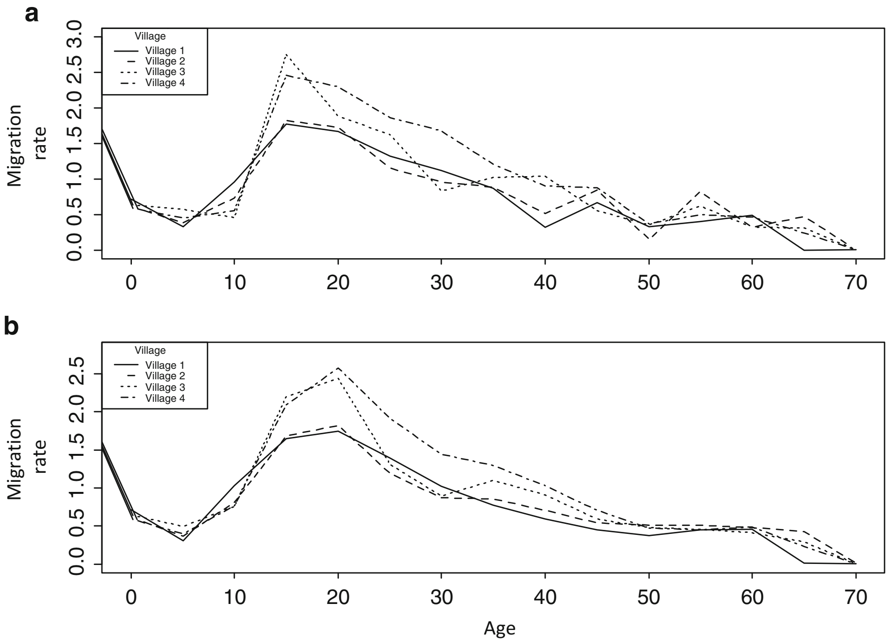

Fig. 16.3 Age-specific migration across the study villages. (**a**) raw data, (**b**) raw data fit to a smoothing spline function with 10 degrees of freedom

Table 16.2 Correlation coefficients (Spearman's and Pearson's) for age-specific migration rates between the study villages

Spearman's	**Village 1**	**Village 2**	**Village 3**	**Village 4**
Village 1				
Village 2	0.86			
Village 3	0.72	0.84		
Village 4	0.86	0.93	0.93	
Pearson's	**Village 1**	**Village 2**	**Village 3**	**Village 4**
Village 1				
Village 2	0.92			
Village 3	0.86	0.91		
Village 4	0.94	0.92	0.93	

(correlograms for each matrix are shown in Fig. 16.5). I then used Mantel tests to look for correlations between locations and out-migration trends. Correlation coefficients and *P-values* were calculated using a Monte-Carlo permutation approach (Tables 16.4 and 16.5)

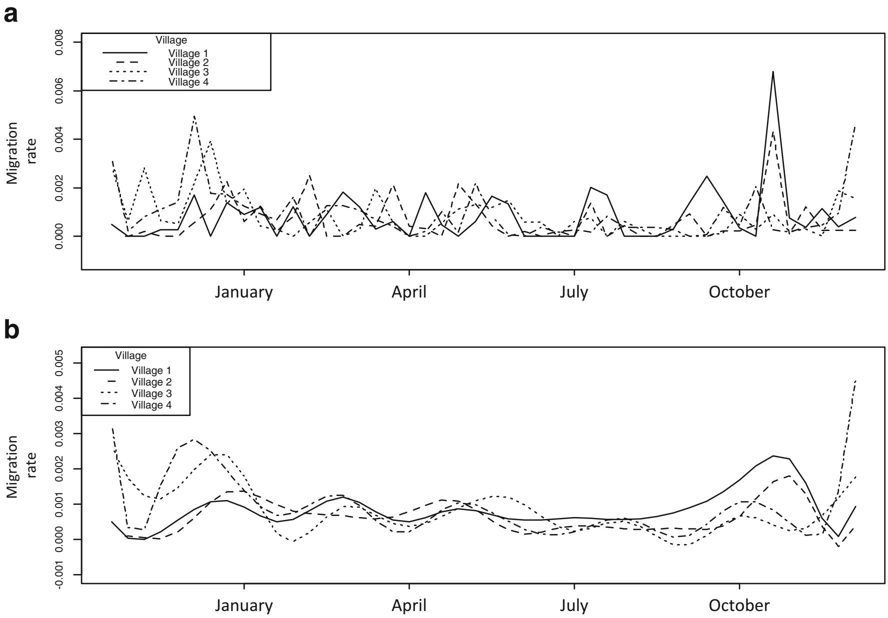

Fig. 16.4 Trends in out-migration in the study villages over the survey period: (**a**) raw data, (**b**) raw data fit to a smoothing spline function with 10 degrees of freedom

Table 16.3 Correlation coefficients (Spearman's and Pearson's) for period-specific migration rates between the study villages

Spearman's	**Village 1**	**Village 2**	**Village 3**	**Village 4**
Village 1				
Village 2	0.17			
Village 3	−0.08	0.16		
Village 4	0.18	0.01	0.36	
Pearson's	**Village 1**	**Village 2**	**Village 3**	**Village 4**
Village 1				
Village 2	0.52			
Village 3	−0.08	0.10		
Village 4	0.03	−0.08	0.48	

16.2.3 Discussion

The age specific patterns in migration indicated in these data are roughly shared with those of lots of other societies (consider the model schedule in Fig. 16.1). Macro-level factors which extent further than the units of study in this brief analysis probably influence these trends in age-specific out-migration. This is interesting for a variety of reasons that I previously alluded to, perhaps especially because it points to law-like patterns in the behavior of human populations. Also, while these

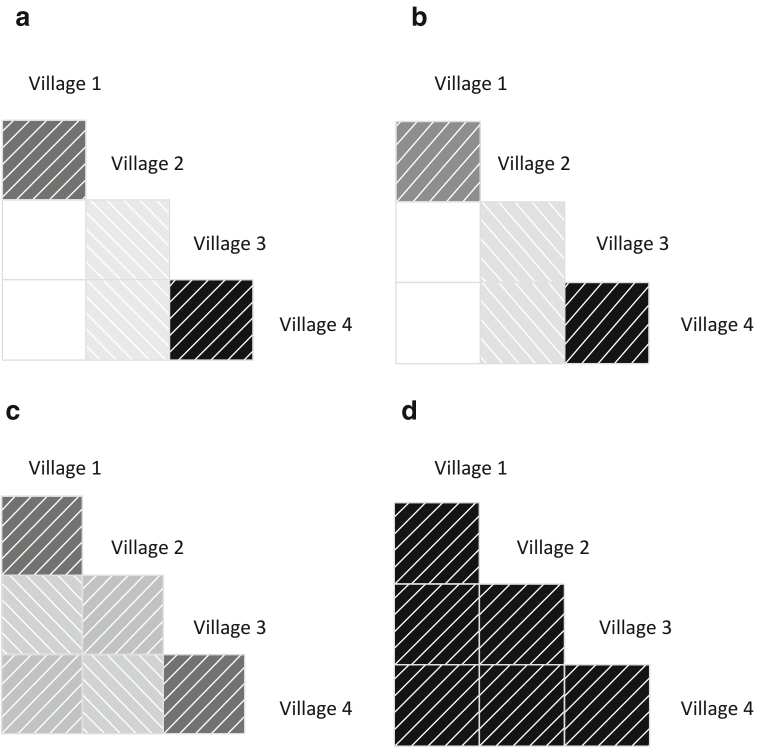

Fig. 16.5 Correlogram indicating: (**a**) distances between villages, (**b**) correlations in elevation between villages, (**c**) correlations in out-migration over time between the study villages, and (**d**) correlations in age-specific out-migration between the study villages *Darker shades* indicate higher correlations, *lines* pointing *up* and to the *right* indicate positive correlations

Table 16.4 Mantel test results for age-specific migration

Matrix		Correlation	*P*-value
Distance	Spearman's	0.7786	0.0001
	Pearson's	0.5021	0.1675
Elevation	Spearman's	0.7281	0.0001
	Pearson's	0.4587	0.2079

Table 16.5 Mantel test results for time series migration

Matrix		Correlation	*P*-value
Distance	Spearman's	0.7924	0.0402
	Pearson's	0.8076	0.1735
Elevation	Spearman's	0.7715	0.0422
	Pearson's	0.7508	0.1660

patterns are shared across so many societies, the components of the migration schedule can generally be broken down into different causal factors for each component of the model schedule. Migration for the very young is likely to occur for different reasons than migration in the elderly. Very young children are

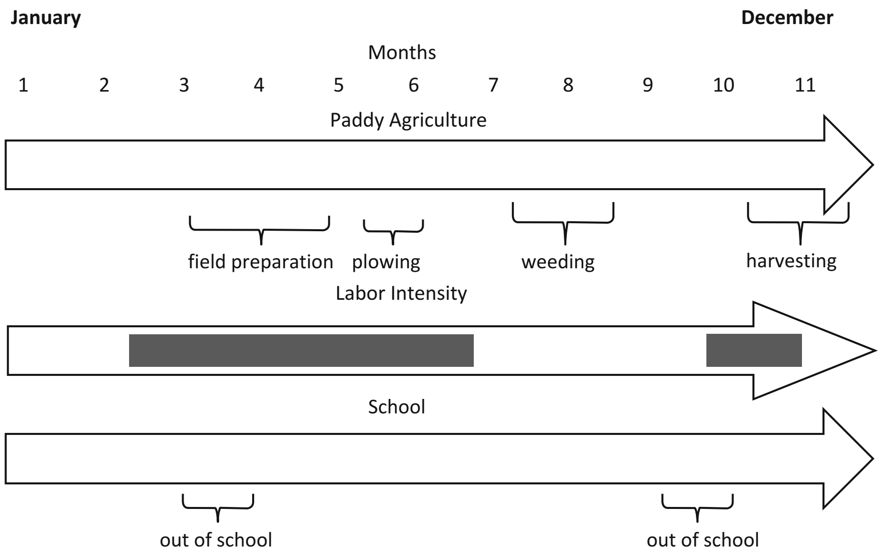

Fig. 16.6 Agricultural and school schedules

typically moving with their parents, school aged children may be moving out to go to school, young adults are likely to move out for marriage or work, and elderly people may move back in with family members or to new places after retirement. My findings with regard to the Karen villages in this study further add credence to the idea that age-specific migration is almost law-like in its shape across the life span, even across vastly different populations.

While the figures illustrating time series migration treat the populations within the villages as homogeneous, the motivations and drivers of such migration may vary. Perhaps the two biggest factors to consider are seasonal migration, which occurs during times of the year when there is little agricultural work, and the timing of the school year for secondary students (Fig. 16.6). Migration which corresponds to the agricultural calendar is hereafter referred to as "seasonal" migration. Such seasonal migration in these populations can mostly be subdivided into two main types: occupational and marital. Seasonal marriage has been noted in many agricultural populations, with peak marriage times occurring during lulls within the agricultural calendar, frequently following the harvests or livestock birthing seasons (Coppa et al. 2001; Gonzalez-Martin 2008; Kussmaul 1985; Wrigley and Schofield 1989). Occupational migration is also likely to occur during these downtimes, with Karen adults sometimes selling their labor or engaging in various forms of trade (Rajah 2008). Conversely, migration of secondary school children out of the household necessarily follows the school calendar (Fig. 16.6). School calendars are roughly coordinated throughout Thailand, meaning that there should be a relatively synchronous out-movement of children as they begin the school semester. Children in rural areas may be more likely than those in urban areas to attend boarding schools and live in or near the school since there are less local schools in those regions. In summary, the rising peak in age-specific migration rates (from around age 12–17), occurs for different reasons (probably schooling) than does migration from 18 until older ages (probably post-marital residence and occupational migration).

While the study villages aren't very far removed from each other geographically, meteorological patterns can vary widely across such relatively small regions. If migration is seasonal, as it appears to be, and if meteorological factors in more proximate villages are more similar than in more distal villages, the patterns shown here may simply coincide with shared macro-level drivers: the Moran effect. One assumption behind the Moran effect is that the population process being observed is density dependent. As previously mentioned, several studies have indicated density dependence in migration patterns. For example, in very small populations there will be a limited amount of potential mates, meaning that marriage aged adults will need to move out in order to find marital partners (Relethford 1986). Conversely, in areas with large populations there may be resource shortages, also potentially leading to out-migration (Umezaki and Ohtsuka 2002). Clearly population density is unlikely to be the *only* important factor in migration, and it may even result in differing dependence relationships in different regions, societies, and cultures (Ohtsuka et al. 1985). However, with regard to the Moran effect, we might expect that different populations with the same density dependencies will experience synchrony in population processes (in our case, migration) when a common macro-level factor effects those same populations. While not confirmatory of density dependence, in the data presented here the largest peaks in age-specific migration are seen in villages with the largest populations (3 and 4, see Table 16.1).

Finally, villages 3, 4, and at least part of 2 share a common secondary school in another village further south of village 4. Children attending secondary school in village 1 also move out to live in dorms when school is in session, but they go to different school(s). There may be variations in academic calendars in these different schools, though those differences are unlikely to be large.

16.3 Scale (in Both Space and Time)

It has been argued that issues of scale, specifically scale and pattern, are *the* "central problem of ecology" (Levin 1992). I would argue that it is also at least *a* central problem in demography, not just spatial demography. While I have already briefly mentioned issues of scale, especially with regard to macro-level drivers in spatial synchrony, the issue isn't limited to space and definitely warrants further discussion. Issues of scale with regard to spatial analysis have been known for some time. For example, ecological analyses have famously been poor predictors of individual outcomes (the ecological fallacy) and the spatial unit that is chosen in a study will almost certainly influence the outcomes of that study (the modifiable areal unit problem) (Robinson 1950; Openshaw 1984).

Scale is also important in temporal dynamics, and the choice of scale can likewise influence analytical results. Some population processes, for example, occur at temporal scales much larger than even typical longitudinal studies (let alone cross sectional studies). Evolutionary changes in populations and population adaptations to climate changes via resettlement and changes in subsistence strategies are but a few examples. Some processes occur at the temporal scale of multiple generations or even longer, rather than a more typical demographic analysis temporal units such as years. Other processes occur in very short temporal units. As previously mentioned, rush hour traffic is the byproduct of short term, quite predictable, human movements. A study on such short term human movements would be impossible if data were collected or analyzed at larger temporal units. Likewise, what happens hourly may not have as much relevance for evolutionary change in human populations.

This concept isn't new to demography. For example, Lotka showed that populations will ultimately reach a stable equilibrium growth rate over time. However, Lotka (and Coale) were also concerned with short term dynamics, and were leery of focusing solely on long-term equilibria, especially when considering demographic dynamics that occur in shorter wavelengths (Lotka 1998; Coale 1972; Caswell 2007). Likewise, applied ecologists have had similar concerns in that the dynamic behavior frequently encountered in nature or over a person's lifespan has less to do with long term (asymptotic) dynamics than with short term (transient) dynamics (Koons et al. 2005; Ezard et al. 2010). Sensitivity and perturbation analyses, along with tests of ergodicity, have indicated that while asymptotic dynamics tend to be insensitive to initial conditions, transient dynamics are very much influenced by the initial state (Caswell and Werner 1978; Caswell 2007). In short, what is important in long term dynamics might have less importance in short term dynamics, and vice versa.

Clearly, scale is important for both spatial and temporal analyses. But how do we know which scale, or since we are considering spatio-temporal analysis, which scales, are the correct ones to use? There is no simple answer. In the previous study I chose villages as my units of analysis, assuming that each village is a separate population. I have also used bi-weekly migration rates on the one hand, and age-specific rates[8] on the other, as temporal units. Other analysts might choose to lump all villages within the region, and others still might instead use the entire district, province, or even nation as their spatial unit. Many demographic analyses conveniently use the year as the basic unit of time.

Probably, the choice of spatial and temporal units should be informed by both the study question at hand and theoretical knowledge. In my case, I am interested in potential synchronous behavior across four study villages, at relatively fine-tuned moments in time. Other researchers might instead be interested in larger units and larger populations and processes that occur over longer periods of time. At least

[8] With 5-year age groups in order to control for potential age-heaping in self reported ages among survey respondents.

theoretically, a prior understanding of the scale at which processes occur should aid in deciding at which scale that process is investigated. It is also important to consider that fine scale data can potentially be aggregated. It may be impossible to adequately disaggregate data collected at very long intervals.

Take, for example, a hypothetical situation in which household wealth directly influences household member out-migration. If a researcher looked at daily household spending and earning she might find ambivalent results whereas if the researcher were to look at monthly or yearly savings she might find a strong, dose-response-like association. This is because the effects of wealth aren't generally instantaneous, they happen over a period of time. Perhaps it is important to consider here that the daily information could potentially be aggregated up to the appropriate temporal scale of months or years but disaggregation is unlikely to work as well. Theory would probably tell us a priori that the effect of wealth (or a lack of it) on the chances that a person migrates is a process that would probably take months or more before leading to an actual migration result. In the absence of a theoretical response, a related approach would be to specifically test for the effects of different time scales – with the end result being theory building or improving.

Conversely, there are methods and techniques that have been designed for defining "natural" populations. Wombling, for example, is a technique that is sometimes used to help define natural boundaries via changes in traits such as allele frequencies in populations (Womble 1951). Briefly, a bilinear function is applied to lattice data in order to create a smooth surface and to calculate vector gradients or surface slopes. Wombling has scarcely been applied to demographic research (but see Bocquet-Appel and Jakobi (1996)). However, it has been proposed that wombling could be used with survival data, with wait times as the quantitative trait of interest: "wombling for wait times[9]".

I leave it to the reader and researchers to decide whether it is more appropriate to define a population *a priori* or via direct methods such as wombling. Such a decision will depend, once again, on the research question at hand. As wombling is a method for defining "natural" borders, and therefore populations, it will necessarily identify relatively homogeneous populations. If heterogeneity within a population is actually what is of interest, then it might not be an appropriate approach.

Regardless, the spatial unit that is chosen is critical for research design and will influence analysis outcomes. A real problem in spatio-temporal analyses, and perhaps especially in demography, is that disaggregated units like the ones used in this chapter are frequently unavailable. Furthermore, the choice of spatial and temporal unit is sometimes driven not by theory or research question but instead by census tracts, zip codes, largely arbitrary district, county, and provincial borders, etc. (Matthews 2011). Countless large scale demographic analyses are flawed by the

[9] Ottar Bjørnstad gave a talk on "wombling for wait times" at an NSF meeting on "Challenges in Modeling the Spatial and Temporal Dimensions of the Ecology of Infectious Diseases" at Ohio State University (Sep. 18, 2012).

inability of demographers to draw data at the correct spatial (and temporal) unit. Conversely, there are real limits to the ability to collect fine-scaled data on large populations, meaning that such data for the types of populations that most demographers enjoy studying are simply unavailable. I'm not sure that these problems can be assumed or ignored away.

16.4 New Data

16.4.1 Computational Considerations with "Big Data"

Research technologies such as global positioning system (GPS) units have drastically changed the ways in which spatial science can potentially be done. Whereas we once had to either assume that people remained confined to their associated spaces and places, or rely on participant memories of where they were at certain times, we can now literally track individual space-time movements. This ability means many things. Perhaps foremost among these is that we can collect more accurate data concerning space-time movements, at spatio-temporal resolutions that were previously impossible. We can now quite literally study hourly (or less) movement patterns.

There are quite a few recent examples of this type of data and research. For example, Pontzer et al. used wearable GPS devices (arm bands) with the Hadza of Northern Tanzania in order to compare the activity patterns of a "traditional" society and Western societies (Pontzer et al. 2012). Quite a few researchers have begun using mobile phone data in studies concerning human movement patterns (Palmer et al. 2012; Raento et al. 2009). For example, Wang et al. used records from cell phone towers in the U.S. to track the movements of individuals and found that daily movements are quite predictable, generally following a power law distribution (Wang et al. 2011). Phithakkitnukoon et al used mobile phone data from Portugal to look at mobility patterns with regard to social networks (Phithakkitnukoon et al. 2012).

But these types of data also come with a cost, not just financially but also in their bulk. That is, "big data" can be quite cumbersome to deal with. It requires new ways of storing data, new ways of analyzing data, new ways of programming, and to some extent, new ways of thinking about the research process. Ten years ago standard personal computers came with hundreds of megabytes of hard disk storage space and today they come with gigabytes; computers with terabytes of hard disk space are becoming increasingly available. Yet despite this incredible growth in computational ability, dataset sizes for many researchers have already outgrown "normal" computer storage capabilities. Furthermore, being able to analyze such data in a timely fashion can be lacking, but technology is catching up to this task as well. High performance computers and institutions with the infrastructure necessary to afford, house, and maintain such networks are a likely turning point for many researchers with such "big data".

Corporations such as Amazon, Rackspace, Microsoft, and Google are increasingly offering their resources, that is, space in the "cloud", for data storage (with petabytes of data) and manipulation. Cloud computing allows users to access, for a fee, supercomputers with excellent capabilities, from virtually anywhere with internet access, rather than having to build their own supercomputers. For example, Amazon's Elastic Compute Cloud service will allow a user to boot "instances" or virtual computing sessions, with a wide range of capabilities, software, and operating systems (Amazon EC2 2013). However, issues surrounding usability, privacy, and the use of confidential data with cloud computing have been voiced (Armbrust et al. 2010).

Storage concerns aside, large data sets are also difficult to work with from a programming and analysis viewpoint. Researchers who do their own data management and statistical coding may need to learn new approaches, software, and coding languages. While computers, memory, and data storage space have become cheaper over time, central processing units (CPUs) have not advanced much with regard to their individual performance. Therefore it is quite common to now see multi-core (multiple CPUs) processors even in personal and office computers, even smart phones. As an extension of the multi-core computer, many researchers have turned to parallel or distributed computing, running statistical models and analyses across multiple computers simultaneously.[10] These computers may be housed in a single location or spread across many locations. Parallel processing allows models or jobs that would once take days to run to be completed in hours. On the other hand, some of that extra time will be spent learning new coding techniques and procedures and likely more overall lines of code that need to be written. Tools such as those available from the Apache Hadoop software library are available for dealing with large data sets and are free of charge (Apache Hadoop 2013; White 2015).

Working across multiple nodes (computers) simultaneously (in parallel) also means that analysts will need to think differently about computational approaches. For example, statistical models in which sequential steps depend on prior steps, where serial processing is necessary, may not benefit from running in parallel. Conversely, if analyses can be divided into multiple components which aren't interdependent, those components can be sent to separate computers, processed in parallel, and then returned to a single core. This type of parallel job is sometimes referred to as being embarrassingly parallel, meaning that it is completely (embarrassingly) separable. Models such as MapReduce, popularized by Google but now available as open source through Hadoop, make the parallel processing approach much more accessible to non-specialists (Dean and Ghemawat 2008). The MapReduce paradigm can basically be divided into two components: the "Map"

[10] For example, a "Beowulf cluster" is a group of mainstream computers that are typically linked together in a central location (rather than spread across many different locations). Beowulf clusters offer a relatively cheap approach to high performance computing.

step and the "Reduce" step. The Map step takes the input and divides it into smaller jobs that are sent to worker nodes whereas the Reduce step collects the output from worker nodes into a final, overarching output at the master node. Most standard statistical software packages are now capable of dealing with large datasets and running in parallel. R ("parallel" and "snow"), SAS ("MP CONNECT"), and Stata ("parallel") all allow users to access and use multiple cores or nodes for various processes.

Furthermore, some high performance computational centers have begun turning to graphics processing units (GPUs) rather than CPUs. Driven in part by demand for better visual capabilities and the gaming industry, GPUs have advanced past the capabilities of CPUs with regard to processing large amounts of data, especially data that are parallel in structure. Whereas CPUs are currently better at dealing with a variety of tasks and data forms, GPUs are exceptional at dealing with large amounts of data and running a few tasks over and over. NVIDIA (2013), a company that manufactures GPUs, has also created a parallel programming platform called CUDA (Compute Unified Device Architecture) which allows users to access computer GPUs and treat them, in some ways, as if they were CPUs NVIDIA (2013).

16.4.2 Ethical Considerations with "Big Data"

New kinds of data mean new kinds of potential ethical issues and there are real ethical issues with regard to collecting such fine scale data. Once again, I draw on the topic of migration as an example. In many cases, migrants and highly mobile people are on the margins of society. They are frequently poor, persecuted, and this may even be a driving factor in their migration. Also, many movements are clandestine or even illegal. For example, in the geographic region where I work, along the Thai-Myanmar border, there is a lot of illegal movement across international borders. As previously noted, much of this movement has been in avoidance of bad circumstances in Myanmar. But that is far from the only reason that people move across this border region. It is also a region that has historically been heavily involved in the illegal trade of narcotics, natural resources and forest products, videos, DVDs and music, livestock, and humans.

A resulting question, then, is: If I could talk people into allowing me to do so, would it be ethical to strap GPS units on individuals and track their movements? Perhaps some better questions are: Could I guarantee that the resulting data won't fall into the hands of government and other agencies that would also very much like to know about movement patterns in this area? Could I ensure my research subjects that no harm will befall them from this type of research?[11] I don't think that I could.

[11] I should note here that the IRB standard is minimal risk to research subjects, weighted against the potential gains from the research. I leave it to individual researchers to decide if that goes far enough and the answer to that would necessarily vary by research topic and research subject. Official ethics policy is an extremely important topic to consider with regard to these new forms of data.

On the other hand, however, knowledge about clandestine movements that are related to human trafficking might be extremely valuable and beneficial. Similar problems and scenarios play out in many places in the world; consider the American Southwest. While the potential for collecting new forms of bigger data is growing, it may be that we sometimes need to collect fewer, more selective data than what is actually possible.

16.5 Conclusions

GPS and big data are both increasingly available to researchers. Arguably, both have the possibility of strongly influencing the future landscape of demography and spatial demographic research. Data that were once only time stamped may now also have spatial references, meaning that spatial relationships can be analyzed in themselves or that unobserved spatial heterogeneity in processes can at least partially be controlled. Furthermore, increases in data availability mean that research projects and topics that were once almost impossible are now increasingly feasible. An obvious example mentioned earlier concerns near real-time human movement patterns which can now be mapped using GPS units or cell phones.

Yet, the availability of such data also presents hurdles that must first be overcome. Chief among these hurdles are the technological and computational burdens associated with extremely large datasets. These technological issues are quickly being addressed and in my opinion, will not remain a significant issue for long. However, demographers who intend to use big data will almost certainly need to learn new approaches to data management, wrangling, and analysis.

Furthermore, big data doesn't only have technological problems; there are tangible theoretical issues as well. Perhaps most of these issues are related to the nature of secondary data use in research. Demographers are probably more comfortable with using secondary data sources than are many researchers in other fields, but issues of data quality, hidden biases, and operationalization remain. Also, having more data isn't always a cure for problems that are inherent in research questions or design. But with the hype surrounding big data, it is quite possible for researchers to think it is such a cure. A carefully designed study, with a small but properly collected and representative dataset, will probably always be superior to the use of extremely bulky, biased, secondary data for ad hoc research questions – no matter how big the size of the latter dataset.

Along with new data collection methods and ultimately new datasets comes the need for new developments in ethics and policy. Current ethics policies are likely to be based on research scenarios from the past (Howell and Porter 2010). Some issues remain the same, but the vast changes in the potential amount and variety of data collected now make confidentiality and safety in research today much different. While ethics policies first need to be updated to deal with modern realities, perhaps a framework that inspires such policies to continually be updated, as research approaches and data capabilities change, should also be considered. Changes in technology and research are now occurring so quickly that by the time ethics and policy are changed to address these changes, it will likely already be time to change them again.

These issues aside, new forms of data also provide new opportunities, not just for addressing theoretical issues but also for creating new methods. Not that long ago, true spatio-temporal analysis, that is, statistical approaches that equally dealt with space and time, were virtually non-existent. Most space-time analyses addressed either space or time explicitly, and at the expense of the other dimension. This is no longer the case – methods for dealing with space and time simultaneously have been proposed, are increasingly available in standard statistical software programs, and continue to develop and emerge. Spatial synchrony, introduced earlier in this chapter, is just one example. Wavelet analysis and wombling are other approaches that are quite relevant to space-time modeling.

Aside from borrowing techniques used in other disciplines, the time is also ripe for the emergence of new methods for analyzing large space and time datasets. Here, with a lot of cumulative experience as data scientists, spatial demographers could have a big role to play. Software packages such as R, which are free and open source, make experimenting and operationalizing new analytical approaches quite approachable, even for researchers with little training in computational sciences. Ultimately, for ambitious and inspired researchers and methodologists with an eye toward the future and with something to offer, the horizon appears to be wide open.

Acknowledgements This study is partially supported by NIAID, NIH (U19AI089672). I would also like to acknowledge data collection by staff at the Vivax Research Center, Mahidol University, the Vector Borne Disease Training Center in Saraburi and staff from the Department of Public Health in Tha Song Yang District, Thailand. Finally, Stephen Matthews, Ottar Bjørnstad and James Wood offered advice with regard to the statistical analysis.

References

Abraído-Lanza, F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the "salmon bias" and healthy migrant hypotheses. *American Journal of Public Health, 89*(10), 1543–1548.

Allee, W. C., & Bowen, E. S. (1932). Studies in animal aggregations: Mass protection against colloidal silver among goldfishes. *The Journal of Experimental Zoology, 61*(2), 185–207.

Amazon Elastic Compute Cloud (Amazon EC2 http). (2013). http://aws.amazon.com/ec2/

Apache Hadoop. (2013). http://hadoop.apache.org/

Armbrust, B., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., et al. (2010). A view of cloud computing. *Communications of the Association for Computing Machinery, 53*, 50–58.

Bjørnstad, O. N., & Bolker, B. (2000). Canonical functions for dispersal-induced synchrony. *Proceedings. Biological sciences/The Royal Society, 267*(1454), 1787–1794.

Bjørnstad, O. N., Ims, R. A., & Lambin, X. (1999). Spatial population dynamics: Analyzing patterns and processes of population synchrony. *Trends in Ecology & Evolution, 14*(11), 427–432.

Bocquet-Appel, J. P., & Jakobi, L. (1996). Barriers to the spatial diffusion for the demographic transition in western Europe. In J. P. Bocquet-Appel, D. Courgeau, & D. Pumain (Eds.), *Spatial analysis of biodemographic data* (pp. 117–129). Paris: John Liberty Eurotext/INED.

Burch, T. K. (2003). Demography in a new key: A theory of population theory. *Demographic Research, 9*(11), 263–284.

Burch, T. K. (2011). Does demography need differential equations? *Canadian Studies in Population, 38*(1–2), 151–164.

Caswell, H. (2006). *Matrix population models*. 2th Edition. Sunderland: Sinauer Associates, Inc.

Caswell, H. (2007). Sensitivity analysis of transient population dynamics. *Ecology Letters, 10*(1), 1–15.

Caswell, H., & Werner, P. A. (1978). Transient behavior and life history analysis of teasel (*Dipsacus Sylvestris* Huds). *Ecology, 59*(1), 53–66.

Cazelles, B., & Stone, L. (2003). Detection of imperfect population synchrony in an uncertain world. *Journal of Animal Ecology, 72*(6), 953–968.

Coale, A. J. (1971). Age patterns of marriage. *Population Studies, 25*, 193–214.

Coale, A. J. (1972). *The growth and structure of human populations*. Princeton: Princeton University Press.

Coleman, D. A., & Haskey, J. C. (1986). Marital distance and its geographical in England and Wales, 1979. *Transactions of the Institute of British Geographers, 11*(3), 337–355.

Coppa, A., Di Donato, L., Vecchi, F., & Danubio, M. E. (2001). Seasonality of marriages and ecological contexts in rural communities of central-southern Italy (Abruzzo), 1500–1871. *Collegium Antropologicum, 25*, 403–412.

Cummings, D. A. T., Irizarry, R. A., Huang, N. E., Endy, T. P., Nisalak, A., et al. (2004). Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature, 427* (6972), 344–347.

Curran, S. R., & Saguy, A. C. (2001). Migration and cultural change: A role for gender and social networks. *Journal of International Women's Studies, 2*(3), 54–77.

De Jong, G. F., & Gardner, R. W. (1981). *Migration decision making: Multidisciplinary approaches to microlevel studies in developed and developing countries*. New York: Pergamon Press.

De Jong, G. F., Richter, K., & Isarabhakdi, P. (1996). Gender, values, and intentions to move in rural Thailand. *International Migration Review, 30*(3), 748–770.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the Association for Computing Machinery, 51*(1), 107–113.

Ezard, T. H. G., Bullock, J. M., Dalgleish, H. J., Millon, A., Pelletier, F., et al. (2010). Matrix models for a changeable world: The importance of transient dynamics in population management. *Journal of Applied Ecology, 47*(3), 515–523.

Fields, G. S. (1976). Labor force migration, unemployment and job turnover. *The Review of Economics and Statistics, 58*(4), 407–415.

Fix, A. G. (1999). *Migration and colonization in human microevolution*. Cambridge: Cambridge University Press.

Frank, R., & Heuveline, P. (2005). A cross-over in Mexican and Mexican-American fertility rates. *Demographic Research, 12*(4), 77–104.

Gonzalez-Martin, A. (2008). Ecological and cultural pressure on marriage seasonality in the principality of Andorra. *Journal of Biosocial Science, 40*, 1–18.

Grenfell, B. T., Bjørnstad, O. N., & Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature, 414*(6865), 716–723.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association, 24*, 7–21.

Harrison, G. A. (1995). Movement and ancestry structure. In G. A. Harrison (Ed.), *The human biology of the English village* (pp. 42–62). New York: Oxford University Press.

Henry, L. (1961). Some data on natural fertility. *Eugenics Quarterly, 8*, 81–91.

Howell, F. M. (1981). Residential preferences and life plans. *Youth & Society, 12*(3), 351–377.

Howell, F. M., & Frese, W. (1983). Size of place, residential preferences and the life cycle: How people come to like where they live. *American Sociological Review, 48*(4), 569–580.
Howell, F. M., & Porter, J. R. (2010). Surveys and geographic information systems. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research*. Bingley: Emerald Publishing Group Limited.
Hudson, P. J., & Cattadori, I. M. (1999). The Moran effect: A cause of population synchrony. *Trends in Ecology & Evolution, 14*(1), 1–2.
Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography*. New York: Springer.
Koons, D. N., Grand, J. B., Zinner, B., & Rockwell, R. F. (2005). Transient population dynamics: Relations to life history and initial population state. *Ecological Modelling, 185*(2–4), 283–297.
Kunstadter, P. (1972). Demography, ecology, social structure and settlement patterns. In A. Boyce & G. Harrison (Eds.), *The structure of human populations*. Oxford: Clarendon.
Kunstadter, P. (1983). Karen agro-forestry: Processes, functions, and implications for socio-economic, demographic, and environmental change in northern Thailand. *Mountain Research and Development, 3*(4), 326–337.
Kussmaul, A. (1985). Time and space, hoofs and grain: The seasonality of marriage in England. *Journal of Interdisciplinary History, 15*, 755–779.
Lee, T. J., Mullany, L. C., Richards, A. K., Kuiper, H. K., Maung, C., et al. (2006). Mortality rates in conflict zones in Karen, Karenni, and Mon states in eastern Burma. *Tropical Medicine & International Health, 11*(7), 1119–1127.
Levin, S. A. (1992). The problem of pattern and scale in ecology: The Robert H. MacArthur award lecture. *Ecology, 73*(6), 1943–1967.
Lieberman, D. (1993). The rise and fall of seasonal mobility among hunter-gatherers. *Current Anthropology, 34*, 599–641.
Liebhold, A., Koenig, W. D., & Bjørnstad, O. N. (2004). Spatial synchrony in population dynamics. *Annual Review of Ecology, Evolution, and Systematics, 35*(1), 467–490.
Lintner, B. (2000). *Burma in revolt: Opium and insurgency since 1948*. Bangkok: Silkworm Books.
Lotka, A. J. (1998). *Analytical theory of biological populations* (Translated and with an Introduction by D.P. Smith and H. Rossert). New York: Plenum Press.
MacDonald, D. H., & Hewlett, B. S. (1999). Reproductive interests and forager mobility. *Current Anthropology, 40*(4), 501–523.
Malthus, T. R. (1817). *An essay on the principle of population, as it affects the future improvement of society, with remarks on the speculations of Mr. Godwin, M. Condorcet, and other writers*.
Matthews, S. A. (2011). Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods. In L. M. Burton (Ed.), *Communities, neighborhoods, and health: Expanding the boundaries of place* (pp. 35–55). New York: Springer.
Matthews, S. A., & Parker, D. M. (2013). Progress in spatial demography. *Demographic Research, 28*(10), 271–312.
Mielke, J. H., Relethford, J. H., & Eriksson, A. W. (1994). Temporal trends in migration in the Åland Islands: Effects of population size and geographic distance. *Human Biology, 66*(3), 399–410.
Moran, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle. *II Synchronization and meteorology. Australian Journal of Zoology, 1*, 291–298.
Muth, R. F. (1971). Migration: Chicken or egg? *Southern Economic Journal, 37*, 295–306.
NVIDIA. (2013). http://www.nvidia.com/content/global/global.php
Ohtsuka, R., Kawabe, T., Inaoka, T., Akimichi, T., Suzuki, T., et al. (1985). Inter- and intra-population migration of the Gidra in Lowland Papua: A population-ecological analysis. *Human Biology, 57*(1), 33–45.
Openshaw, S. (1984). *The modifiable areal unit problem. Concepts and techniques in modern geography* (Vol. 38). Norwich: Geo Books.

Palmer, J. R. B., Espenshade, T. J., Bartumeus, F., Chung, C. Y., Ozgencil, N. E., et al. (2012). New approaches to human mobility: Using mobile phones for demographic research. *Demography, 50*(3), 1105–1128.

Parrado, E. A., & Morgan, S. P. (2008). Intergenerational fertility among hispanic women: New evidence of immigrant assimilation. *Demography, 45*(3), 651–671.

Phithakkitnukoon, S., Smoreda, Z., & Olivier, P. (2012). Socio-geography of human mobility: A study using longitudinal mobile phone data. *PLoS ONE, 7*(6), e39253.

Pontzer, H., Raichlen, D. A., Wood, B. M., Mabulla, A. Z. P., Racette, S. B., et al. (2012). Hunter-gatherer energetics and human obesity. *PLoS ONE, 7*(7), e40503.

Porter, J. R., & Howell, F. M. (2012). *Geographic sociology: Theoretical foundations and methodological applications in the sociology of location* (GeoJournal library, Vol. 105). Dordrecht/New York: Springer.

Portes, A., & Sensenbrenner, J. (1993). Embeddedness and immigration: Notes on the social determinants of economic action. *The American Journal of Sociology, 98*(6), 1320–1350.

Pred, A. (1977). The choreography of existence: Comments on Hägerstrand's time-geography and its usefulness. *Economic Geography, 53*(2), 207–221.

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods & Research, 37*(3), 426–454.

Rajah, A. (2008). *Remaining Karen: A study of cultural reproduction and the maintenance of identity*. Acton: The Australia National University.

Ranta, E., Kaitala, V., Lindstrom, J., & Linden, H. (1995). Synchrony in population dynamics. *Proceedings: Biological Sciences, 262*(1364), 113–118.

Raymer, J., & Rogers, A. (2008). Applying model migration schedules to represent age-specific migration flows. In J. Raymer & F. Willekens (Eds.), *International migration in Europe: Data, models and estimates* (pp. 175–192). Chichester: Wiley.

Relethford, J. H. (1986). Density-dependent migration and human population structure in historical Massachusetts. *American Journal of Physical Anthropology, 69*, 377–388.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*(3), 351–357.

Rogers, A., & Castro, L. J. (1981). *Model migration schedules*. Laxenburg: International Institute for Applied Systems Analysis.

Romaniuk, A. (2011). A comment on Thomas K. Burch's paper "Does demography need differential equations?". *Canadian Studies in Population, 38*(1–2), 165–167.

Royama, T. (1992). *Analytical population dynamics*. London/New York: Chapman & Hall.

Stark, O., & Bloom, D. E. (1985). The new economics of labor migration. *American Economic Review, 75*(2), 173–178.

Stark, J. H., Cummings, D. T., Ermentrout, B., Ostroff, S., Sharma, R., et al. (2012). Local variations in spatial synchrony of influenza epidemics. *PLoS ONE, 7*(8), e43528.

Swedlund, A. C. (2009). Mating distance and historical population structure: A review. In C. G. N. Mascie-Taylor & A. J. Boyce (Eds.), *Human mating patterns* (pp. 15–30). Cambridge: Cambridge University Press.

T. B. C. (2004). *Between worlds: Twenty years on the border*. http://www.refworld.org/docid/4cb2c3ee2.html

Thompson, W. S. (1929). Population. *American Journal of Sociology, 34*, 959–975.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.

Todaro, M. P. (1980). *Internal migration in developing countries: A survey. Population and economic change in developing countries* (pp. 361–402). Chicago: University of Chicago Press.

Umezaki, M., & Ohtsuka, R. (2002). Changing migration patterns of the Huli in the Papua New Guinea Highlands. *Mountain Research and Development, 22*(3), 256–262.

Voss, P. R. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*(5), 457–476.

Wachter, K. W. (2005). Spatial demography. *Proceedings of the National Academy of Sciences, 102*(43), 15299–15300.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A. et al. (2011). *Human mobility, social ties, and link prediction categories and subject descriptors*. (pp. 1100–1108) Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

White, T. (2015). *Hadoop: The definitive guide* (4th ed.). Sebastopol: O'Reilly.

Wijsman, E. M., & Cavalli-Sforza, L. L. (1984). Migration and genetic population structure with special reference to humans. *Annual Review of Ecology, Evolution, and Systematics, 15*, 279–301.

Winterhalder, B., & Smith, E. A. (1992). *Evolutionary ecology and human behavior* (Foundations of human behavior). New York: Aldine Transaction.

Womble, W. H. (1951). Differential systematics. *Science, 114*(2961), 315–322.

Wood, J. W., Smouse, P. E., & Long, J. C. (1985). Sex-specific dispersal patterns in two human populations of Highland New Guinea. *The American Naturalist, 125*(6), 747–768.

Wrigley, E., & Schofield, R. (1989). *The population history of England 1541–1871*. Cambridge: Cambridge University Press.

Part IV
Instruction in Spatial Demography and Concluding Remarks

Chapter 17
Instruction in Spatial Demography

Stephen A. Matthews

17.1 Introduction

In December 2011 a specialist meeting on *Future Directions of Spatial Demography* brought together specialists from multiple disciplines to discuss the state of the science in spatial demography, emergent geospatial data and measurement issues, and spatial statistical methods (for further details on this specialist meeting see Matthews et al. 2012).[1] It is not the intent to review and update the discussions that took place at this meeting but rather to focus on arguably the most important cross-cutting theme that emerged from the meeting: *instruction in spatial demography*.

While demography is inherently a spatial science (Weeks 2004) the training many demographers receive in fundamental spatial concepts, geospatial data, and analytical methods is often limited, patchwork, or nonexistent. However many important demographic questions deserve to be studied and framed using spatial approaches and this will become even more evident as changes in the volume, source, and form of available demographic data—much of it geocoded—further changes the data landscape and thus the methods demographers need to use. Ultimately changes in the data demographers collect, how they collect data, how they link data, and how they analyze data suggest the need to train next-generation

[1] The majority of attendees were geographers and sociologists, many other disciplines were represented, including anthropology, economics, epidemiology, health economics, and political science. Most participants were interested in demographic research questions. Full details about the specialist meeting, including a participant list, short position papers from participants, copies of presentations made during the meeting, and additional materials (including this report) can be found at http://ncgia.ucsb.edu/projects/spatial-demography/

S.A. Matthews (✉)
Department of Sociology & Department of Anthropology, State College, PA, USA
e-mail: matthews@psu.edu; sxm27@psu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_17

population scientists in spatial thinking, concepts, and methods of analysis. This will be a challenge given the existing logistical constraints on instruction in spatial demography. That is, any emergent method has to compete for its own place within the established curricula which will include core courses (e.g., fertility, migration, mortality), an expansive range of substantive cores (e.g., population and environment, urbanization) and other methods courses (e.g., demographic techniques, event history analysis, multilevel modeling). In this chapter I will begin by discussing spatial perspectives in demographic research; this is both a selective review and as an introduction to emergent trends in geospatial data and methods. This opening section illustrates a major challenge associated with instruction in spatial demography, namely the breadth of topics that legitimately fall under an umbrella of spatial demography. Next I describe available instructional resources (courses, textbooks, software and other resources), few of which focus on demographic research, and then transition to a discussion of potential new directions and strategies (action items) in instruction in spatial demography.

17.2 Trends in Spatial Perspectives in Demographic Research

If we are to understand the instructional needs in spatial demography we need to understand some of the main emergent trends in the field. Below I identify four main trends.

First, there has been a rapid growth of interest in the addition of a spatial perspective in demographic research. In part, this growth is driven by the ready availability of geospatial data and the refinement and emergence of GIScience tools to analyze them: geographic information systems (GIS), spatial analysis, and spatial statistics (de Smith et al. 2007; Matthews 2011). For some time now, many social sciences have begun to accept spatial analysis as part of their various methodologies and some see the spatial perspective serving as a potential incubator for innovative social science and interdisciplinary research (Goodchild et al. 2000; Goodchild and Janelle 2004, 2010; Janelle and Goodchild 2011). Indeed, in a commentary in *Science*, Butz and Torrey (2006) identified geographic information science (GIScience) as one of six innovative frontiers in social science research.

Second, many research and policy questions faced by demographers require analysis of complex patterns of interrelated social, behavioral, economic, and environmental phenomena (Castro 2007). Increasingly it is argued that spatial thinking and spatial analytical perspectives have important roles to play in confronting such complexity (Voss 2007; Goodchild and Janelle 2010; Porter and Howell 2012). In recent years, the revival in macro-demography has been supplemented by the integration of micro- and macro-demography and the linking of data on people to data on places (Entwisle 2007). Demographic research in the United States (U.S.) and overseas increasingly depends on the collection and analysis of individual- and contextual-level data across a wide range of spatial and temporal scales. Indeed, this is evident in demographic research on such issues as racial/ethnic segregation and other forms of social stratification and inequality; health behaviors, morbidity and mortality; fertility; family structure/transitions and aging; and population-environment interactions (Entwisle 2007, see also several position papers from the specialist meeting, available at: http://ncgia.ucsb.edu/projects/spatial-demography). For example, the inclusion of geocodes in many Demographic and Health Surveys has and will continue to transform the types of research questions that can be asked as researchers begin to integrate micro and macro data by linking individual data to land use and land cover, climate, built environment and other dimensions of contextual environments.[2] However, as Burgert et al. (2012, p. 627) note in a response to Mansour et al. (2012) "using the DHS GPS data for analysis requires a proper understanding of the data collection, displacement, and verification procedures to assess the spatial error in the data and the limitation this may place on using the data at different spatial scales." Other new methods focusing on data integration include the use of remote sensing (e.g., night-time lights databases for looking at urban spatial change and measles [see Bharti et al. 2011, Chen 2016, Chap. 15, this volume]). In summary, there is a continuing need for closer integration between spatial concepts, data, and methods and the enduring micro and macro frameworks of analysis in demography.

Third, the micro-macro integration has piqued the interest of researchers in the harnessing of geospatial technologies to collect, manage, and analyze new forms of geospatial data that can help address research and policy questions. The volume, sources, and forms of geospatial data are growing rapidly. Thus, many well-used demographic data sets and products increasingly include (or retrofit) geocodes and contextual attributes, and some have started to include geographic boundary or shapefiles. In addition, there are also several domestic and international data harmonization projects that not only standardize on attributes, but also on geographical units. These include for example, projects such as those from the Minnesota

[2] While discussions of spatial demography tend to be US focused it is important to acknowledge that international demographic research has long included an explicit spatial perspective to how demographic research is framed and how the data are collected, organized, and analyzed. Liverman et al. (1998) and Fox et al. (2003) both discuss the challenges and opportunities associated with the use of remote sensing, GIS, and spatial econometrics and demonstrate how these tools have been used effectively to analyze the relationship between human activities and local environmental change. More recently, Weeks et al. (2013) illustrate how spatial concepts, data, and methods can be integrated in a study examining spatial inequalities in Accra, Ghana. Indeed, while the review to date has focused on US-based spatial demography innovation in spatial demography in international research has been high. Entwisle et al. (1997) was among the first papers to appear in *Demography* that explicitly used geospatial data (GIS and GPS) and spatial methods (spatial network analysis). GPS data were integrated with survey and administrative records for Nang Rong, Thailand, and this permitted a more nuanced analysis of contraceptive choice. In another early paper, Guilmoto and Rajan (2001) provided a rare illustration of the use of spatial correlograms and kriging methods applied to the study of fertility within India; the spatial variation in fertility across districts in India was not random and the spatial structure of fertility decline had intensified over time.

Population Center, such as the School Attendance Boundary Information System (www.sabindata.org) and National Historic GIS for the U.S. (www.nhgis.org) and the international projects such as Terra Populus (http://www.terrapop.org/) and the Integrated Public Use Microdata Series-International (https://international.ipums.org/). In addition, there are commercial products such as GeoLytics normalized Neighborhood Change Database 1970–2000 and other products (www.geolytics.com).

Developments in participant-generated data or volunteered geographical information (Goodchild 2007) also warrant serious attention, including data from Twitter feeds and traces from global positioning systems (GPS). In addition next-generation wireless and sensor technologies, and new data storage and handling technologies (e.g., cloud computing, geospatial data warehouses, data mining techniques, and relational databases) are already, and will continue to, change what, how, and when we collect data on individuals and their environments. New data formats that tag geographic location and time provide unparalleled spatial and temporal precision. In a world of 'big data' and wide distribution of smart phones, the idea of compiling intensive longitudinal data on individuals has traction (Borell 2011). This is especially the case in studies of movement, networks and interactions (for applications in sociology and demography see González et al. 2008; Raento et al. 2009; State et al. 2013; Palmer et al. 2013).

Fourth, and related to the previous paragraph, in the near future we should anticipate significant changes in how demographers conduct fieldwork as new data collection technologies fundamentally change the quality, scope, and flexibility of measures we collect and use about the social, built, and physical environments. That is, the collection of new types of individual and area-based geospatial data will greatly facilitate the measurement of appropriately defined contexts and individual exposure to contexts. These technological developments and enhancements in new spatio-temporal precision have enormous potential to permit better functional understandings of human spatial behavior, providing new ways of thinking about relative and absolute utilization and/or exposure to place, but also raising important issues about spatial embeddedness and scales of analysis (Chaix et al. 2009; Matthews and Yang 2013; Palmer et al. 2013). In turn this should generate closer links between theory, data, and methods in multilevel analysis of demographic and health outcomes (Entwisle 2007). Similarly, emerging statistical methods and new types of data coupled with reciprocal enhancements in conceptual models will help promote spatially informed demographic research.

Many other challenges exist, the most obvious ones relating to data privacy, data validation, utilizing data on non-representative samples of people and places, data preservation (historic data) and archiving. The increasing availability of precise, accurate spatial data on individuals and their activity paths (via GPS and cell-phone tracking), and the ease with which these data can be integrated with other contextual databases and health data implies that the need for training in handling confidentiality and privacy issues has never been greater. This need was highlighted in the National Research Council report on confidentiality and spatial data (Gutmann and Stern 2007), has been reinforced in papers found in recent special issues on spatial demography (VanWey et al. 2005; Gutmann et al. 2008), and has been echoed in social science and health research more generally (Boulos et al. 2009). The degree to which data privacy concerns are discussed in the current offerings of 'spatial' courses is not known but based on a small sample of online syllabi (see http://gispopsci.org/curriculae/) these topics do not frequently or directly appear as a lecture/seminar topic or in readings. That brings us to the focal issue of this chapter; instruction in spatial demography.

17.3 Instruction in Spatial Demography

In 2002 Menken, Blanc, and Lloyd noted that the broadening of the field of demography would necessitate the acquisition of additional skills and familiarity with the concepts and tools of related disciplines.[3] Spatial analysis is one such area. Indeed, the need to think spatially and use spatial analytical tools has never been higher. Although the focus in spatial demography courses is often on methods, it is important to teach concepts and to have a solid foundation in spatial thinking. If demographers can be convinced that space matters for specific types of demographic questions, they will be interested in enhancing their spatial analytical skill set. Unfortunately, the incorporation of spatial thinking and analysis into a demography training program seems to be more easily said than done.

The instruction in fundamental spatial concepts, geospatial data, and analytical methods in demography training programs is often limited, patchwork or nonexistent. To provide context to this rather dire pronouncement it is important to examine existing instructional materials, specifically *courses* and *textbooks*. And, as additional context it is important to note that even within academic settings where the in-house instructional capacity and infrastructure (i.e., trained instructors and available resources) is in place, traditional models of training in the discipline of demography often pose logistical constraints. That is, any emergent instructional foci—such as spatial demography—has to compete for attention within the established demography curricula; which typically includes core courses (e.g., on fertility, migration, mortality), an expansive range of substantive courses (e.g., family formation, population and environment, urbanization), and established methods courses (e.g., demographic techniques, event history analysis, multilevel modeling). That is, even for those programs that can offer a graduate-level spatial demography course, there are challenges in trying to fit the course within the wider curricula at a specific institution (Matthews et al. 2012). The instructor will most likely be able to develop a one semester/quarter course. In this scenario, concerns exist over how much depth an instructor can and should provide regarding micro-macro demography or whether to focus on specific spatial statistical methods. There

[3] Among the recommendations from 'Rethinking the teaching of demography,' Palloni (2002, p. 57) included a brief sentence suggesting that our training programs include "an option to learn about the nature and application of spatial statistics."

are many fundamental questions. *What to teach? How to teach?* And, *How to integrate the spatial and the demographic?* It is important to note that spatial analysis incorporates a wide variety of techniques and, thus, even *What to teach?* becomes problematic due to the wide array of methods that might be introduced and the depth at which they are covered.[4] Moreover, it is highly unlikely that one instructor can keep up with all of the new and existing data and methods. Inevitably, the potential exists for much foundational material (basic concepts) to be left out at the expense of focusing on a 'method', though even here what is included in the syllabus will depend on factors such as prerequisite courses. Thus, even in the traditional instructional environment it can be a challenge to accommodate and promote training in spatial demography. Compromises inevitably must be made.

17.3.1 Courses

As has been implied the field of spatial demography is changing rapidly. Unfortunately while the instructional environment for introductory GIS and spatial analysis courses continue to improve, the application and use of spatial theory and advanced spatial analysis methods in demography appears to lag other fields (Matthews et al. 2007, 2012). This lag is not helped by the limited availability of graduate-level training in geospatial data, concepts, cartography, and GIS and more specifically, the lack of courses in advanced spatial data analysis—such as spatial regression modeling or spatial pattern analysis—with a significant social science or demographic content. Indeed, several position statements at a 2011 specialist meeting on *Future Directions in Spatial Demography* alluded to an observation that most of the universities that host a population center do not host a geography department nor do they have the capacity for a sustained instructional program on spatial thinking and spatial analysis (e.g., see position papers by Castro, Curtis, Howell, Pan, Rey, Riosmena, and Voss among others at http://ncgia.ucsb.edu/projects/spatial-demography/).[5] Although some universities offer classes in software (e.g., ArcGIS), very few specifically teach spatial thinking and spatial analytical methods.[6] This is a problem. Not only is the lack of spatial training in demography a

[4] The diversity of the student body seeking out spatial demography courses is also a challenge. While there is high demand for such courses, often the diversity of the students' background and experience (in geospatial data handling, cartography, quantitative methods) and their substantive interests creates other challenges associated with how to pitch and introduce spatial analytical methods.

[5] It is important to note that even at universities with geography programs and/or a GIS capacity, spatial training—especially training in advanced spatial analysis methods—may not be associated with geography and GIS programs.

[6] Paradoxically, there has been a growing number of online GIS certificate and Masters programs (see http://ucgis.org/gis-degree-programs) and professional organizations such as the University Consortium for Geographic Information Science (UCGIS)—www.ucgis.org—developed model

challenge for the field, but the fact that very few 'demographers' receive any training at all in spatial methods, or exposure to spatial thinking, prior to entering graduate school compounds the issue. Judging from parallel situations in other fields that have recently embraced GIS and spatial analysis, such as public health, preventive medicine, and epidemiology (Matthews 2012), the implication is quite clear. We need to re-think how to provide training in spatial demography to early-career demographers at institutions without formal courses.[7]

17.3.2 Textbooks

The provision of GIS and spatial analysis courses is but one dimension of training. What about the inclusion of spatial thinking and spatial analytic methods in demography textbooks? Within standard demography methods texts, the treatment afforded spatial analysis is scant to say the least (see Hinde 1998; Preston et al. 2001; Siegel and Swanson 2004). Indeed, among the best known demographic techniques and methods texts the references to spatial analysis are indirect and typically arise in sections discussing, for example, the geographical hierarchies of census data or national and sub-national (i.e., regional) attribute data. Lacking in many textbooks is the discussion of foundational concepts in spatial thinking (Janelle and Goodchild 2011; and see www.teachspatial.org). This is quite a serious instructional gap for a field that is an 'inherently spatial science' (Weeks 2004, p. 381). Weeks' 11th edition of *Population* (2011) is a rare exception among upper-division demography texts in that it includes modest coverage of GIS, geospatial data, and spatial analysis. Another rare exception is Namboodiri (1991) who included a chapter titled 'Spatial Distribution,' that introduces spatial data and methods—spatial probability models and point pattern analysis, spatial autocorrelation, spatial regression, and methods for dealing with correlated error terms—that could be applied to demography. Currently no specialist text on what one may regard as "spatial demography" exists, with the nearest equivalent arguably found in applied population geography texts (e.g., Plane and Rogerson 1994) and demography-related fields such as the book on *GIS and Public Health* by Cromley and McLafferty (2002; second edition, 2011) and a book for political scientists (Darmofal 2015). I note there have been publications on spatial population analysis (Rees and Wilson 1977; Woods and Rees 1986) and multiregional demography (Rogers 1975, 1995).

GIS curricula (UCGIS 2006) and is currently updating their *Geographic Information Science and Technology Body of Knowledge* 2015 Project.

[7] It is worth noting that while the emphasis here is on graduate instruction in the US, the attendees of the specialist meeting recognized the need for training in spatial demography at *all* education levels (pre-university, undergraduate, graduate, and postgraduate) as well as infrastructure and initiatives that could facilitate exposure to spatial thinking and analysis across more distributed scholars, internationally.

If demography textbooks are scant on GIS and spatial analysis the reverse is also often true, that GIS and spatial analysis texts rarely cover demographic topic areas. An encouraging recent development is the appearance in introductory GIS textbooks and workbooks of more attention to applied social, health, and demographic issues. However, in these introductory texts the treatment of spatial *analysis* beyond cartography, spatial querying, overlay, and buffer analysis is either non-existent or minimal; and this is especially so of, otherwise useful guidebooks and tutorials specific to commercial software products (e.g., Kurland and Gorr 2012). Moreover, many tutorial textbooks or workbooks typically offer up a sanitized GIS experience unlike the real world. At the other end of the spatial analysis textbook market there are the advanced spatial statistics texts; Cressie's (1991) classic text on *Statistics for Spatial Data* immediately comes to mind (and Cressie and Wikle 2011). More specifically, there are several spatial econometrics texts (Anselin 1988; LeSage and Pace 2009). In these high-end texts there is little focus on demographic research applications and also this is largely true of the well regarded intermediate texts (Bailey and Gatrell 1996; Haining 2003; O'Sullivan and Unwin 2010), primers (Fotheringham et al. 2002; Bivand et al. 2008), and handbooks (Anselin and Rey 2010; Fischer and Getis 2010; Nyerges et al. 2011; Fotheringham and Rogerson 2009). An emergent trend in textbooks is the online (interactive) text. In the area of spatial analysis the most comprehensive available is de Smith et al. (2013) (www.spatialanalysisonline.com).

Supplementing university courses and textbooks are two other modes of instruction: *workshops and online resources*, including the use of web-based instruction. There are several excellent sites providing multiple resources related to spatial analysis. Excluding the main commercial vendors among the very best academic institution-based site is the GeoDa Center at Arizona State (http://geodacenter.asu.edu). This provides a treasure trove of instructional materials focused around GeoDa software and PySAL, an open source library of computational tools for spatial analysis. Among the more general purpose sites include the Center for Integrated Spatial Social Science (www.csiss.org now archived) and the Center for Spatial Studies (http://spatial.ucsb.edu) both at the University of California, Santa Barbara (UCSB). The latter site is one of the most comprehensive in the social sciences providing information on learning resources, spatial resources, spatial tools, events, and literature searches. The TeachSpatial (http://teachspatial.org/) site offers a comprehensive set of resources for 'spatial learning and teaching,' focusing on fundamental spatial concepts. A GIS and Population Science website (www.gispopsci.org)—a collaboration between Penn State and UC Santa Barbara—is tailored more specifically to population scientists; and is based on materials and resources compiled for workshops focusing on spatial demography.[8]

[8] Several centers and organizations in the US and overseas will offer workshops on spatial analytic methods. These are usually one off events typically lasting 1 week. Similarly, spatial analytic methods are occasionally the subject of pre-conference workshops for national and international conferences; including the Population Association of America.

Although it is possible to find training opportunities and resources to learn advanced spatial analysis methods in the university sector, the commercial sector, and from textbooks, the opportunities are limited, costly, and frequently not targeted towards population science research questions and applications. One message from the *Future Directions of Spatial Demography* specialist meeting is that the instructional resources there is a need to employ multiple and diverse, but integrated forms of instructional delivery; i.e., including workshops, self-paced instructional materials, and to take advantage of webinars.

17.4 Future Strategies to Enhance Instruction in Spatial Demography . . . and Raise the Visibility of the Field

While there are opportunities for instruction in spatial demography these tend to be concentrated in a very small number of academic institutions and research centers. In anticipation of the changing availability and use of geospatial data by demographers this instructional model is no longer sustainable. Indeed, it would be prudent to explore the use of instructional models that allow for the pooling of experts and resources across centers (e.g., webinars and online resources) and invest in approaches that will substantially increase the number of instructors (i.e., *train the trainers* via focused workshops). Below I briefly outline three action items that could help increase the number of next generation instructors, promote awareness of spatial theories, concepts, data, methods, and applications, all with the goal to enhance the quality of demographic scholarship. If there is to be a sea-change in spatial demography instruction then what is proposed below should be integrated activities or action items not stand alone. It is noted that making any resources available in multiple languages (e.g., Spanish, French) would be a huge stride forward in helping promote spatial informed demographic research in many parts of the world.

Action Item #1: Develop a model curricula and textbooks in spatial demography

While the visibility of spatial demography is enhanced by a focused journal—*Spatial Demography* (www.spatialdemography.org)—and special issues of journals a more ambitious undertaking would the development of a model curricula and spatial demography textbooks. A useful service to a new researcher in this area would be one that coordinates the collection of syllabi, provides guidance to graduates on what to learn and master vis-à-vis geospatial data products, software, and methods.[9] Unfortunately for the new researcher the field of spatial statistical analysis covers a wide collection of methods, each specific to different forms of geospatial data they might use (i.e., depending on whether the unit of analysis is represented within a spatial analysis model as a point, a line, a polygon, a grid, or a

[9] The new journal, *Spatial Demography*, provides regular reviews of data, code, and software.

surface). These methods include geostatistics, pattern analysis, exploratory spatial data analysis, spatial econometrics, geographically weighted regression, multilevel modeling, remote sensing and image analysis as well as topics such as population forecasting, small area estimation, spatial sampling, spatial uncertainty, and methods to mask or protect against data disclosure. In sum, spatial demography can draw on all of these methods and this can create problems vis-à-vis course design, instructional priorities, and workshop content; an issue further compounded by diversity of interests and abilities among potential audiences. As noted earlier few established spatial demography courses do exist (several syllabi are available at www.gispopsci.org). While model curricula have emerged for GIS instruction, broadly defined, nothing comparable exists for training demographers in spatial analysis. A model curricula would potentially consist of materials that would comprise multiple semester long courses or short courses not just one ‘generic’ course; that is, the model curricula would potentially include a course or sections of courses that would emphasis training in fundamental spatial concepts (e.g., scale, distance, units of analysis, spatial heterogeneity, spatial dependence), or on cartography and geovisualization, on exploratory spatial data analysis, and so on. Ideally, a model curricula could be designed to be modular or flexible allowing ‘instructors’ to piece together material for a specific course offering.

Action Item #2: Train the trainers

A conventional and successful approach to learning has been the workshop or short course model.[10] Workshops that focus on ‘*training the trainers*’ (i.e., the practice of teaching spatial demography) could potentially have a large and sustained impact on the field. The goal of a *training the trainers* workshop should be to assist faculty, postdocs, and graduate students (future instructors) in creating opportunities for their own students to engage in critical spatial thinking about demographic issues.[11] I would submit that early systematic exposure to spatial thinking and analysis will enhance student interest in pursuing graduate studies, research, and careers that draw on their exposure to the concepts, tools, and applications of spatial demography, using different types of data. Substantively,

[10] The workshop format can be successful for both general training in spatial thinking and introductory spatial methods (see http://csiss.ncgia.ucsb.edu/GISPopSci/) and also in advanced methods workshops on specific methods such as spatial econometrics, pattern analysis, and geographically weighted regression (see http://csiss.ncgia.ucsb.edu/GISPopSci/). Matthews was PI of the grant that offered these workshops.

[11] Ideally workshop participants would review methods of instruction (e.g., use of open-source software, project-based exercises, classroom communication, and peer interaction). Participants would have opportunities to leverage teaching innovations (e.g., syllabi, course demonstrations, exercises, and learning assessments) by sharing their creations through a website, and this could serve the dual purpose of helping to build up collections of resources that would easily enable instructors to embed spatial thinking within their own demography curricula. This kind of model was used by CSISS in their Space workshops series (see http://www.spatial.ucsb.edu/affiliates/space.php).

the focus of such workshops must be on integrating demographic theories with spatial analysis and the geo-visualization and interpretation of demographic data.

While workshops can be highly successful, if there are to be significant impacts on training in spatial demography we need to embrace new technologies to supplement these learning opportunities, and we need to provide learning infrastructure resources that can both help support instruction and more importantly facilitate a network of scholars with shared substantive and analytical interests from across demographic and related research disciplines. This leads to…

Action Item #3: Webinars

An emerging training model is predicated on the idea that there is a need for an expanded range of instructional delivery methods based on webinars.[12] I use the term webinars to define any instructional event involving both instructors and participants distributed across two or more sites. Webinars have emerged as a common forum for meetings, seminars and instruction. At a basic level they enable distributed individuals (e.g., in the field of spatial demography) to meet together in a forum to share and discuss topics. The webinar format can allow users to pool resources across countries, academic institutions, and population agencies as well as couple substantive and methodological experts. Rather than one off webinar events it would be arguably more beneficial to have coordinated series offered for multiple levels of expertise and interest (i.e., webinars that address topics from fundamental spatial concepts, or focus on new technologies and data, or on advanced analytical techniques). Webinars can take on many formats; that is, while some webinars might be lecture-based, others could interactive and include panel discussions. Increasingly webinars are events that are supported by other materials (e.g., reading materials, lecture notes, labs, logs of questions submitted and responses). Activities like webinars and related discussion forum can also help generate a network of spatial demography scholars across the globe. Indeed, potentially training resources can be generated (and updated) by a few centers and experts to be pooled for use by anyone at any demographic institution anywhere in the world. In this way, webinars can supplement in-house training programs and thus provide a unique translational service across the field of demography more generally.

These action items are unlikely to emerge without the active engagement of the spatial demography community.

[12] Webinars are already used by U.S. federal agencies (e.g., U.S. Census Bureau—http://www.census.gov/mso/www/training/training_events.html) and many others besides.

17.5 Discussion

It is important to remind ourselves of the most salient demographic topics of the twenty-first century—migration, urbanization, health, aging, vulnerable populations, and inequality—and in doing so consider our theoretical frameworks, our data needs, and the methods we will likely use. Spatial demographers will play an important role in addressing the salient topics as they are concerned with the study of demographic processes and outcomes in a way that admits the effects of space and place on individual life-course trajectories, the spatial-temporal nature of an individuals' exposure to risk of demographic events, and the complexity this introduces into statistical analysis and the visualization of demographic data. A spatial perspective on demographic processes and outcomes thus includes the use and adoption of spatial concepts, geospatial data, spatial technologies, and spatial analytic methods.[13] Changes in the data demographers collect, how they collect data, how they link data, and how they analyze data suggest a critical need to train the next-generation of population scientists in spatial thinking, concepts, and methods of analysis.

The future of spatial demography is bright but we also need to ensure that instructional opportunities do not lag too far behind the new data and methods encountered by demographers. In this chapter I raised several concerns about the instructional gap. The instructional gap already exists but without action will potentially widen. We are not training the next-generation of demographers to think spatially, to embrace new and emerging geocoded data, or to use spatial analytical techniques. Individual researchers will inevitably be able to secure the necessary training for their own interests and needs but access to appropriate instruction in fundamental concepts and basic forms of analysis is not routinely available at the undergraduate level, let alone the graduate level. When we extend our gaze to colleagues in developing countries the access to experts and instructional resources is even more restricted. As a consequence I argue that there is a need to re-think how to provide training in spatial demography. Specifically, I suggest the need to explore approaches that can reach diverse and widely distributed audiences as well as capture the advantages of both synchronous and asynchronous learning. To have a sustained impact on instruction I suggest that we need to explore how we can *'seed and populate'* the next generation of 'teachers' and facilitate the development and introduction of new courses (or sections of courses) at both the graduate and undergraduate levels. We need to be more creative with the use of webinar events, and continue to develop new and maintain existing on-line resource guides and materials. Paralleling the development of instructional resources, and possibly emergent from such activities would be the creation of model curricula and textbooks in spatial demography. The latter would raise the visibility of spatial demography.

---

[13] Studying and thinking about a spatial world does not always translate to a study of place, and as such spatial demography remains distinct from population geography.

References

Anselin, L. (1988). *Spatial econometrics, methods, and models*. Dordrecht: Kluwer.
Anselin, L., & Rey, S. (2010). *Perspectives on spatial data analysis*. Berlin: Springer.
Bailey, T. C., & Gatrell, A. C. (1996). *Interactive spatial data analysis*. Harlow: Longman.
Bharti, N., Tatem, A. J., Ferrari, M. J., Grais, R. F., Djibo, A., & Grenfell, B. T. (2011). Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science, 334*(6061), 1424–1427.
Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer.
Borrell, B. (2011). Every bite you take. *Nature, 470*, 320–322.
Boulos, M. N. K., Curtis, A. J., & AbdelMalik, P. (2009). Musings on privacy issues in health research involving disaggregate geographic data about individuals. *International Journal of Health Geographics, 8*, 46.
Burgert, C. R., Zachary, B., & Way, A. (2012). Response to "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*, 626–627.
Butz, W., & Torrey, B. B. (2006). Some frontiers in social science. *Science, 312*, 1898–1900.
Castro, M. (2007). Spatial demography: An opportunity to improve policy making at diverse decision levels. *Population Research and Policy Review, 265*, 477–509.
Chaix, B., Merlo, J., Evans, D., Leal, C., & Havard, S. (2009). Neighborhoods in eco-epidemiologic research: Delimiting personal exposure areas: A response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine, 69*, 1306–1310.
Chen, X. (2016). Using nighttime lights data as a proxy in social scientific research. In F.M. Howell et al. (Eds.), *Recapturing space: New middle-range theory in spatial demography*. Cham: Springer.
Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley.
Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Hoboken: Wiley.
Cromley, E. K., & McLafferty, S. L. (2002). *GIS and public health*. New York: Guilford Press.
Cromley, E. K., & McLafferty, S. L. (2011). *GIS and public health* (2nd ed.). New York: Guilford Press.
Darmofal, D. (2015). *Spatial analysis for the social scientist*. Cambridge: Cambridge University Press (David Darmofal's book will appear in October 2015 (see http://www.cambridge.org/US/academic/subjects/politicsinternational- relations/politics-general-interest/spatial-analysis-social-sciences?format=PB)).
De Smith, M. J., Goodchild, M. F., & Longley, P. (2007). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools*. Leicester: Troubador Publishing Ltd.
De Smith, M. J., Goodchild, M. F., & Longley, P. A. (2013). *Geospatial analysis* (4th ed.). Leicester: The Winchelsea Press, Troubador Publishing Limited. Available at http://spatialanalysisonline.com/
Entwisle, B. (2007). Putting people into place. *Demography, 444*, 687–703.
Entwisle, B., Rindfuss, R. R., Walsh, S. J., Evans, T. P., & Curran, S. R. (1997). Geographic information systems, spatial network analysis, and contraceptive choice. *Demography, 34*(2), 171–187.
Fischer, M. M., & Getis, A. (Eds.). (2010). *Handbook of applied spatial analysis: Software tools, methods and applications*. Berlin: Springer.
Fotheringham, A. S., & Rogerson, P. A. (Eds.). (2009). *The SAGE handbook of spatial analysis*. Thousand Oaks: Sage.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.
Fox, J., Rindfuss, R. R., Walsh, S. J., & Mishra, V. (Eds.). (2003). *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*. Norwell: Kluwer Academic Publications.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*, 779–782.
Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.
Goodchild, M. F., & Janelle, D. G. (Eds.). (2004). *Spatially integrated social science*. New York: Oxford University Press.
Goodchild, M. F., & Janelle, D. G. (2010). Toward critical spatial thinking in the social sciences and humanities. *GeoJournal, 75*, 3–13.
Goodchild, M. F., Anselin, L., Applebaum, R. P., & Herr Harthorn, B. (2000). Toward spatially integrated social science. *International Regional Science Review, 23*, 139–159.
Guilmoto, C. Z., & Irudaya Rajan, S. (2001). Spatial patterns of fertility transition in Indian districts. *Population and Development Review, 27*(4), 713–738.
Gutmann, M. P., & Stern, P. C. (Eds.). (2007). *Putting people on the map: Protecting confidentiality with linked social-spatial data*. Washington, DC: National Academies Press.
Gutmann, M. P., Witkowski, K., Colyer, C., McFarland O'Rourke, J., & McNally, J. (2008). Providing spatial data for secondary analysis: Issues and current practices relating to confidentiality. *Special Issue: Spatial Demography. Population Research and Policy Review, 27*(6), 639–665.
Haining, R. (2003). *Spatial data analysis: Theory and practice*. Cambridge: Cambridge University Press.
Hinde, A. (1998). *Demographic methods*. London: Arnold.
Janelle, D. G., & Goodchild, M. F. (2011). Concepts, principles, tools and challenges in spatially integrated social science. In T. L. Nyerges, H. Couclelis, & R. McMaster (Eds.), *The SAGE handbook of GIS and society* (pp. 27–45). Thousand Oaks: Sage.
Kurland, K. S., & Gorr, W. L. (2012). *GIS tutorial for health* (4th ed.). Redlands: ESRI Press.
LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC.
Liverman, D., Moran, E. F., Rindfuss, R. R., & Stern, P. C. (Eds.). (1998). *People and pixels: Linking remote sensing and social science*. Washington, DC: National Academy Press.
Mansour, S., Martin, D., & Wright, J. (2012). Problems of spatial linkage of a geo-referenced Demographic and Health Surveys (DHS) dataset to a population census: A case study of Egypt. *Computers, Environment and Urban Systems, 36*, 350–358.
Matthews, S. A. (2011). Spatial analysis. In *Oxford bibliographies online—Sociology*. http://www.oxfordbibliographiesonline.com/view/document/obo-9780199756384/obo-9780199756384-0058.xml
Matthews, S. A. (2012). Thinking about place, spatial behavior, and spatial processes in childhood obesity. *American Journal of Preventive Medicine, 42*(5), 516–520.
Matthews, S. A., & Yang, T.-C. (2013). Spatial Polygamy and Contextual Exposures (SPACEs): Promoting activity space approaches in research on place and health. *American Behavioral Scientist, 57*(8), 1057–1081.
Matthews, S. A., Janelle, D. G., & Goodchild, M. F. (2007). *Advanced spatial analysis training for population scientists*. R25 NICHD grant: http://gispopsci.org/prior-awards. (Note that the grant was written in 2007 and awarded in 2008)
Matthews, S. A., Janelle, D. G., & Goodchild, M. F. (2012). *Future directions in spatial demography specialist meeting: Final report* http://ncgia.ucsb.edu/projects/spatial-demography/docs/Future-Directions-in-Spatial-Demography-Report.pdf
Menken, J., Blanc, A. K., & Lloyd, C. B. (Eds.). (2002). *Training and support of developing-country population scientists: A panel report*. New York: Population Council.
Namboodiri, K. (1991). *Demographic analysis: A stochastic approach*. San Diego: Academic Press Inc. Harcourt Brace Jovanovich, Publishers.
Nyerges, T. L., Couclelis, H., & McMaster, R. (Eds.). (2011). *The SAGE handbook of GIS and society*. Thousand Oaks: Sage.
O'Sullivan, D., & Unwin, D. J. (2010). *Geographic information analysis* (2nd ed.). Hoboken: Wiley.
Palloni, A. (2002). Rethinking the teaching of demography: New challenges and opportunities. *Genus, LVIII*(3–4), 35–70.

Palmer, J. R., Espenshade, T. J., Bartumeus, F., Chung, C. Y., Ozgencil, N. E., & Li, K. (2013). New approaches to human mobility: Using mobile phones for demographic research. *Demography, 50*(3), 1105–1128.

Plane, D. A., & Rogerson, P. A. (1994). *The geographical analysis of population: With applications to planning and business*. New York: Wiley.

Porter, J. R., & Howell, F. M. (2012). *Geographical sociology: Theoretical foundations and methodological applications in the sociology of location*. Dordrecht: Springer.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Malden: Blackwell.

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods and Research, 37*, 426–454.

Rees, P. H., & Wilson, A. G. (1977). *Spatial population analysis*. London: Edward Arnold.

Rogers, A. (1975). *Introduction to multiregional mathematical demography*. New York: Wiley.

Rogers, A. (1995). *Multiregional demography: Principles, methods and extensions*. New York: Wiley.

Siegel, J. S., & Swanson, D. A. (2004). *The methods and materials of demography* (2nd ed.). San Diego: Elsevier Academic Press.

State B., Weber, I., & Zagheni, E. (2013, February). Studying inter-national mobility through IP geolocation. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 265–274). ACM.

UCGIS. (2006). *Geographic information science and technology book of knowledge* (1st ed.). Washington, DC: Association of American Geographers.

VanWey, L. K., Rindfuss, R. R., Gutmann, M. P., Entwisle, B., & Balk, D. L. (2005). Confidentiality and spatially explicit data: Concerns and challenges. *Special Issue: Spatial Demography. Proceedings of the National Academy of Science, 102*(43), 15337–15342.

Voss, P. (2007). Demography as a spatial social science. *Population Research and Policy Review, 26*(5), 457–476.

Weeks, J. (2004). The role of spatial analysis in demographic research. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science* (pp. 381–399). New York: Oxford University Press.

Weeks, J. (2011). *Population* (11th ed.). Belmont: Wadsworth/Thomson Learning.

Weeks, J., Hill, A. G., & Stoler, J. (Eds.). (2013). *Spatial inequalities: Health, poverty, and place in Accra, Ghana*. Dordrecht: Springer.

Woods, R., & Rees, P. H. (Eds.). (1986). *Population structures and models: Developments in spatial demography*. London: G. Allen and Unwin.

Chapter 18
Concluding Remarks: Developing Spatial Demography

Frank M. Howell, Jeremy R. Porter, and Stephen A. Matthews

The goal of this book has been to advance thinking in the specialty of spatial demography through enhancing middle range theory. While inevitably an incomplete statement, the authors contributing chapters have pushed what we know forward in some measure. Perhaps more importantly is the potential contribution of this edited volume in providing a set of chapters with a common goal of addressing traditional demographic issues within the framework of spatial analysis. In this final chapter, we summarize the major thrust of each chapter by section. We then emphasize some key themes arising from these results with a critical view for what else is needed. The next section focuses on what we need to know in the short run for spatial demography to strengthen its position as a scientific enterprise. Finally, we articulate some directions for the near future for spatial demography.

First, we should spend some time re-examining the contributions made by the authors of the chapters in this volume. As a collection, these chapters examine current issues associated with spatial thinking, introduce methods associated with the conceptualization of space, link spatial thinking to established literature across a number of disciplines, test theory using spatial methods, operationalize "place" based on theory, and push forward our current measurement and examination of contemporary topics in the social sciences through the introduction of new methods

F.M. Howell
Department of Sociology, Mississippi State University, 416 Bay Cove,
Ridgeland, MS 39157, USA
e-mail: frankmhowell@gmail.com

J.R. Porter (✉)
Brooklyn College and Graduate Center, City University of New York, Brooklyn, NY, USA
e-mail: jporter@brooklyn.cuny.edu

S.A. Matthews
Department of Sociology & Department of Anthropology, State College, PA, USA
e-mail: matthews@psu.edu; sxm27@psu.edu

F.M. Howell et al. (eds.), *Recapturing Space: New Middle-Range Theory in Spatial Demography*, Spatial Demography Book Series 1,
DOI 10.1007/978-3-319-22810-5_18

and measures from outside disciplines. In sum, the goal of this collection of chapters was to continue the tremendous work others in the area of spatial analysis have established. Most directly related to this goal, Chap. 2 (Logan) in this volume discusses the *Challenges of Spatial Thinking* discusses the link between theory and method in area of spatial analysis.

Siordia and Matthews (Chap. 3) continue to push this idea by linking the conceptualization of place to the theoretical justification for spatial analysis. Directly related to that, Brazil (Chap. 4) discusses the inherent effect of place conceptualization and operationalization as it pertains to the meaning of the concept of "neighborhood effects". Taken together, these two chapters underlie the current issues associated with much of spatial analysis as it currently stands. On the one hand, we are measuring something that has yet to be satisfactorily conceptualized and on the other hand, we are reporting findings that are related to probabilistic statistical associations related to spatial processes, but we don't really understand the mechanisms through which they operate. Wong (Chap. 5) and Weeks (Chap. 6) close out the first part of this volume by discussing the application of evolving spatial methods and theory in a way that push forward their interconnectedness and their utility in any research that explicitly incorporates a spatial approach.

The second part of this volume is dedicated to research in practice and its extension to understanding important issues that are inherently spatial but have not been treated as such. Mobley and Bazzoli (Chap. 7) extend much of the literature in the area of hospital choice and increase the utility of its modeling through the incorporation of spatial dimensions. Darmofal and Strickler (Chap. 8) continue this theme by extending influential research in Political Science with a spatial approach. In both cases the utility of spatial analysis in areas where it has yet to be applied proves to highlight the important contribution spatial approaches can make to our understanding of what we currently believe are well established and understood areas of disciplinary research.

Part III of this volume is made up of the Chaps. 9, 10, 11, 12, 13, 14, 15, and 16 which focus in the areas of theory development and testing. Chapter 9 (Shin and Agnew) test the link between voter turnout and spatial considerations in Italy and find that understanding turnout more accurately requires attention focused on local context. Chapter 10 (Porter and Howell) focus on the development of a more theoretically informed geography for the capture of population redistribution across the rural-urban continuum, Chap. 11 (Irwin and Pischke) use theoretically informed spatial gravity models to examin hydro-fracturing activity in Pennsylvania, and Chap. 12 (Tolbert, Blanchard, Mencken, and Li) examine civic community in regards to variation across time and space. All three of these (Chaps. 10, 11, and 12) focus on the measurement of community per the theoretically appropriate catchment areas associated with the substantive topic being studied.

Chapter 13 (Yang, Shoff, and Noah) and Chap. 14 (Kramer) both focus on the methodological development of spatial methods through added sophistication. The former re-examines the rural-urban paradox by identifying theoretical and methodological issues associated with measurement of place and the ability to incorporate explicitly exogenous and endogenous relationships with mortality using a spatial Durbin model. The latter examines the variations in effects associated with the measurement of "local" and how that relates to our understanding of effects on pregnancy outcomes. The final two chapters of Part III, Chap. 15 (Chen) and Chap. 16 (Parker), both push forward the boundaries of data and methods used by social scientists through the use of satellite data and spatial synchrony models. In both cases, the ability to find "new" and "better" proxies for issues/concepts of interest to social scientists incorporating spatial methods is a necessary step in moving forward our ability to understand a number of substantive topics that are currently examined with unsatisfactory "state-of-the-art" methods. Finally, Chap. 17 (Matthews) discusses the incorporation and transmission of concepts, methods, and theory associated with spatial analysis through the process of instruction.

18.1 What Do We Know About Spatial Theory/Methods and Their Applications?

At a high level of abstraction, here are some guiding ideas about the state of theory in spatial demography as this chapter is being written.

18.1.1 Space Is Important … But We Are Less Sure About Why

A growing number of social researchers have uttered the mantra that "space matters" (in some cases very influentially, for examples see Voss 2007; and Tickamayer 2000) without identifying much more than grand theory as to why it does. Others (e.g., Lobao et al. 2007), argue similarly but do point out the need for interstitial work on the middle ground. As the debate over the famous Tobler's Law in geography has concluded, the dictim that "closer" things are more alike than things "further" away, is more of an organizing principle than an actual law (Tobler 2004; Goodchild 2004). Indeed, Barnes (2004) goes further: "Adding law talk, though, does not contribute anything to it substantively; it is like paying the work an empty compliment. Instead, my claim is that to understand and to appreciate fully that substance, we need to examine specific local practices." (Barnes 2004: 283). This immediately brings to mind work on residential segregation, itself which has garnered quite an empirical consideration in the spatial measurement and conceptual aspects of the concept. Only recently has Wong, Chap. 5 (this volume) and others pushed forward the theoretical underpinnings of this concept in spatial terms (also see Lee et al. 2008; Reardon et al. 2008; Osth et al. 2014).

In many ways, work to date has followed the Law of the Instrument: "Give a small boy a hammer, and he will find that everything he encounters needs pounding." (Kaplan 1964). This trend is not unusual or unexpected. The technology leading up to the so-called "spatial revolution" (see Chap. 1) simply gave demographers the technology to "do" spatial analysis. Much like the computing revolution and large scale population surveys with the exemplar of Blau and Duncan's *The American Occupational Structure* (1967) which led social science journals to become festered with path analysis models purporting to identify causal structures and processes, the advances leading to the spatial revolution has fostered much focus on the method with middle range theory lagging far behind. We believe that it is past time to advance theoretical understanding of why spatial processes work rather than rely on the refrain of "space matters".

18.1.2 What Do We Need to Know?

The "laws" of Tobler, and to a lesser extent, Zipf (1949), have served as the theoretical point(s) of departure for much recent era spatial analysis. That is, those basically conducting research through the application of "new" spatial methods with the expectation that "space matters". But how do we move beyond those types of research designs by building middle range theory? As the exploratory work on spatial inequality by Lobao et al. (2007) began to make the case in a recent edited volume on this topic:

> ...contributors to this volume share a view ... of the similarities among approaches that study inequality at different scales. These similarities include the recognition that analysts are addressing essentially common questions about stratification across scales, building from critically oriented theory, and using comparative methodological approaches. At the same time, within sociology the topic of spatial inequality itself remains unevenly developed.

In turn, Matthews (2012) discussed the obvious similarities between spatial demography and population geography. In doing so, Matthews made the case that the two were actually different in regards to methods as most research in geography text does not use spatial analysis but spatial demography is predicated on the application of spatial methods. This indicates that not only is spatial demography spatial in the sense that geography is often incorporated into analyses, but also explicitly spatial in the approaches used to develop literature in the area. (Collectively, we do note that spatial demography can be based upon spatial theory, independent of spatial methods.)

Other similar arguments can be made for spatial demography but how should that progress be constructed over the next decade? Here are some strategic ideas:

18.1.3 No Spatial Association

Instead of "everything is connected to everything," such as many who have used Tobler's Law have argued, it would facilitate stronger theory to identify those social demographic dimensions that do NOT follow this spatial pattern. In structural equations modeling, for instance, it is as important to find no effect on an outcome as it is a significant causal one. Related more directly to the spatial methods around which this text has been developed, an analogy can be made to better understand the application of spatial diagnostics in regression modeling through the testing and identification of the optimal weighting schemes, the differentiation between the character of the weighting scheme (contiguity, distance, catchment, etc.), and their ability to capture the spatial process in question as a result.

Explicit spatial tools and methods provide a good framework from which we can move forward in these endeavors. For example, the open source software package GeoDa (and the manuals that accompany the program) provide a good framework for moving through the process of understanding diagnostics and correctly handing spatial processes that may affect the reliability of related statistical analyses. In addition, approaches such as Geographically Weighted Regression (GWR) have been used to identify spatial non-stationarity in regression modeling. In both cases, the GeoDa framework and the application of GWR models can be used to confirm the lack of any spatial processes through the application of explicitly spatial methods. In this way, spatial analysis in an exploratory form should become part of traditional research process with the understanding that confirming spatial processes either way is important to producing good science in the field of demography. Understanding and confirming that there is a lack of a spatial process at play could be thought to be just as important as confirming and incorporating its existence.

18.1.4 Umbrella Test

The spatial scale of a spatial analysis is always an issue; that is, what scale is optimal for the phenomenon being investigated? One strategy used by researchers has been to recompute analyses over several spatial scales—such as Census tracts, ZIP codes, and block groups—to see how similar or robust the results are. What do these results show? Whether they are identical (unusual) or wildly different or somewhere in between, does it tell us much about the phenomenon? It depends on how we invoke middle range suppositions about the data. If it's acknowledged as exploratory, the differences by spatial scale may illustrate something worth further study. If it is deductive, then similar results *replicate* while disparate ones *specify*, following the Elaboration Model of survey research (Hyman 1955). But it is critical to acknowledge conceptual and theoretical elements of what spatial scale is most relevant to the phenomenon being studied. The first author has called this the "umbrella test." We get weather forecasts for states every day. But is such a large-scale forecast at all useful for individuals or localities? What smaller scale forecast would be optimal? One that determines whether one may need an umbrella! *The most useful expression of the spatial scale is the desired one*.

The logic underlying the *Umbrella Test* is parallel to the LISA test developed by Anselin (1995) for expanding upon the global Moran's I test for spatial randomness. While it is important to have statistical guidance as to the spatial randomness of the entire extent of the study, that tells us little about the locality(ies) of importance. So it is too regarding the theoretical importance of the spatial scale.

18.1.5 Dissolve Cross-Disciplinary Boundaries

A number of disciplines pursue lines of study complementing spatial demography: economics (including agricultural economics), history, archeology, anthropology, geography, epidemiology, political science, sociology (including rural sociology), statistics, and mathematics are a few. Cross-fertilization from one discipline to another has often resulted in critical exemplars. The importation of path analysis by Duncan into sociology in his collaboration with Blau in the *American Occupational Structure* (1967) is a clear example. By contrast, such an intentional cross-fertilization can be contrasted with the intentional walls that we often maintain. That being said, it is important to note that boundaries between disciplines are dissolving overall. One of the co-authors of this chapter notes that he has "reviewed over 400 grants for NIH and never do sole investigator R01s score well; the innovation is in collaboration and teams—and I don't mean 4 sociologists working together, I mean researchers across different fields working together (e.g., the cutting edge work in population and environment research is often collaborative between demographers/social scientists and environmental scientists)". Similar experiences are likely to be reported by many across a number of different fields and point to the increasingly important place of interdisciplinary areas like spatial demography, which draw together and synthesize information and research from many different disciplines.

One of the clearest examples of this involve the famous Chicago School of Social Ecology (Abbot 1999). As pointed out in *Geographic Sociology* (2012), the founders of this perspective borrowed their perspective and methods from the rural sociologist Charles J. Galpin, one of the first to use map displays to illustrate and study social relations within a county administrative boundary. Park et al. (1925) openly acknowledged this intellectual debt in the *AJS* itself. However, the Chicago group forgot this by the time their most popular works were published and the subsequent historians like Abbot (1999) had well forgotten that lineage. Why is that important? Because Galpin's work is very relevant today for spatial demography but only those trained in rural sociology would know of it. His use of maps to spatially identify the relations that residents of the county had with key community institutions is precisely what leading spatial demographers are attempting to replicate today (see, for example, Logan's research on the Urban Transition Project (http://www.s4.brown.edu/UTP/index.htm). To this point, part of the purpose of this book is to address the issue of fractured literatures and knowledge of previously employed methods in the area of spatial analysis (and spatial demography in particular). Some of the "new" approaches to the spatial analysis of social data are not new at all in either concept or practice. However, the lack of a field-wide understanding of the roots of spatial analysis in the social sciences limits our ability to build on what's been done and what is currently "state-of-the-art". We are not the first to draw attention to this and much larger projects aimed at centralizing such information have been undertaken at places like the Center for Spatially Integrated Social Sciences (CSISS) at UCSB. Our hope is that we can contribute to the development of our foundational knowledge with this volume.

18.1.6 Central Concepts for Middle Range Theory

It could perhaps be said that the building block of spatial demography is the "county" because of the ubiquity of public data associated with that administrative polygon. Indeed, counties do "act" in terms of government and non-government jurisdictional actions and services (e.g., Lobao et al. 2012). However, the practical utility for the county boundary aside, our needs for central core concepts for furthering middle range theory push us toward sub-county geographies. Indeed, the Holy Grail for spatial demographers, sociologists, and many others social scientists interested in studies of groups and organizations (human or not) is the concept of "community." Those demographers without institutional training or affiliation with rural sociology (or agricultural economics) departments tend to universally ignore the century-long and productive history regarding the community concept in rural sociology (see Hayes 1908; Porter and Howell 2012). We need more integrative strategies combining approaches that build upon the key notions for a social ecology of community. Two strands of thinking at present appear in contradistinction to one another: those using extant Census-based geographies ("top down") and those using crowd-sourced behavior or sentiments ("bottom up"). The latter push the social network aspects of "community" while the former emphasize the legalized basis for Census "places" or the more localized and smaller geography of "neighborhood" through the Census tract.

However, others might argue that this is too crude of a distinction as it is generally thought that demographers use area data while, for instance, public health researchers are increasingly making use of point level data. Demographers often assume an individual is 'geocoded' to a tract which then defines a neighborhood. But in physical activity (and even crime research), researchers often draw concentric circles or street network buffers around an individual's home address as a data point. Scholars have different intellectual traditions on which they draw. Many social scientists are lattice or area data analysts while in other fields they are point, line and surface analysts. Spatial demography, as a field of scientific inquiry, has an outstanding opportunity to be more complete in it's corpus of knowledge in the sense that its interdisciplinary nature provides a platform for a constant cross-fertilization of theory and methods if boundary-maintenance remains permeable.

Are these approaches incompatible? We do not believe so. Perhaps once again reflecting the disciplinary blinders afflicting many scholars, one largely under-recognized approach does explicitly combine both aspects. The Kaufman-Wilkinson Field Theory approach (Kaufman 1959; Wilkinson 1970, 1991) has been continued by Luloff and his associates (Luloff and Wilkinson 1977; Theodori and Luloff 2000; Flint et al. 2010). Proponents of this perspective identify core elements of community as "territory, local society or networks of association, and community field or the process of expressing common interests of the local society" (Flint et al. 2010: 28). The territorial-based networks of association involves agency among the coming together of people who recognize local needs. "A key component to community is the generalized bond that emerges when people come together to act because they share spaces and problems (Wilkinson 1991). Community identity is a driving force in promoting local action." (Flint et al. 2010; 32; additionally see Grannis 2009).

We believe that these generalized bonds are often spatially-expressed. The standing commuting vectors from one small area to another may provide key insights into interactional community boundaries (Farmer and Fotheringham 2011; Comber et al. 2012). Standing behavioral flows like work commuting, shopping, service access, and recreation reflect transportation-based vectors while the consumption of locality-based news and advertising are conveyed through contiguous communications markets. The symbolic identity of territorial localities provide a common bond among organizations and their networks. The first author terms these linkages as *spatially-expressed social bonds*. A considered reading of Galpin's bulletin which gave impetus to the Chicago School of Social Ecology (Park et al. 1925) will reveal that these patterns were the key elements of his community institution networks illustrated in his maps of Walworth County, WI. We believe that an effective marriage of Galpin's "social anatomy" approach with the Kaufman-Wilkinson Field Theory of community using contemporary spatial methods would result in significant progress in middle range results regarding the core concept of an ecologically-based community.

References

Abbott, A. (1999). *Department and discipline Chicago sociology at one hundred*. Chicago: University of Chicago Press.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.

Barnes, T. J. (2004). A paper related to everything but more related to local things. *Annals of the Association of American Geographers, 94*(2), 278–283.

Comber, A. J., Brunsdon, C. F., & Farmer, C. J. Q. (2012). Community detection in spatial networks: Inferring land use from a planar graph of land cover objects. *International Journal of Applied Earth Observation and Geoinformation, 18*, 274–282.

Farmer, C. J. Q., & Fotheringham, A. S. (2011). Network-based functional regions. *Environment and Planning A, 43*(11), 2723–2741.

Flint, C. G., Luloff, A. E., & Theodori, G. L. (2010). Extending the concept of community interaction to explore regional community fields. *Journal of Rural Social Sciences, 25*(1), 22–36.

Goodchild, M. F. (2004). GIScience, geography, form, and process. *Annals of the American Association of Geographers, 94*, 709–714.

Grannis, R. (2009). *From the ground up translating geography into community through neighbor networks*. Princeton: Princeton University Press.

Hayes, E. C. (1908). Sociology and psychology; sociology and geography. *American Journal of Sociology, 14*(3, November), 371–407.

Hyman, H. (1955). *Survey design and analysis: Principles, cases, and procedures*. New York: Free Press.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science* (p. 28). San Francisco: Chandler.

Kaufman, H. F. (1959). Toward an interactional conception of community. *Social Forces, 38* (October), 8–17.
Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review, 73*, 766–791.
Lobao, L. M., Hooks, G., & Tickamyer, A. R. (Eds.). (2007). *The sociology of spatial inequality*. Albany: SUNY Press.
Lobao, L., Jeanty, P. W., Partridge, M., & Kraybill, D. (2012). Poverty and place across the United States: Do county governments matter to the distribution of economic disparities? *International Regional Science Review, 35*(2), 158–187.
Luloff, A. E., & Wilkinson, K. P. (1977). Is community alive and well in the inner city? *American Sociological Review, 42*(5), 827–828.
Matthews, S. A. (2012). *Frontiers in spatial demography and population geography*. Presentation given at the annual meetings of the Association of American Geographers, New York.
Osth, J., Clark, W. A. V., & Malmberg, B. (2014). Measuring the scale of segregation using k-nearest neighbor analysis. *Geographical Analysis, 47*(1), 34–49.
Park, R. E., Burgess, E. W., & McKenzie, R. D. (1925). *The city*. Chicago: University of Chicago Press.
Porter, J. R., & Howell, F. M. (2012). *Geographical sociology: Theoretical foundations and methodological applications in the sociology of location* (GeoJournal library, Vol 105). Springer: New York, NY.
Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography, 45* (3), 489–514.
Tickamayer, A. R. (2000). Spatial inequality in the future of sociology. *Contemporary Sociology, 28*(6), 805–813.
Theodori, G. L., & Luloff, A. E. (2000). Urbanization and community attachment in rural areas. *Society & Natural Resources: An International Journal, 13*(5), 399–420.
Tobler, W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers, 94*(2), 304–310.
Voss, P. R. (2007). Demography as a social science. *Population Research and Policy Review, 26*, 457–476.
Wilkinson, K. P. (1970). The community as a social field. *Social Forces, 48*(3), 311–322.
Wilkinson, K. P. (1991). *The community in rural America*. Westport: Greenwood Press.
Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.