UNIVERSITY OF AMSTERDAM

Incorporating Space in Multimethod Research: Combining Spatial Analysis with Case-Study Research

Harbers, I.; Ingram, M.C.

Incorporating Space in Multimethod Research: Combining Spatial Analysis with Case-Study Research

Imke Harbers, *University of Amsterdam*
Matthew C. Ingram, *University at Albany, SUNY*

Most outcomes that social scientists care about—including democracy, development, institutions, participation, and violence—are not distributed randomly across geographic space. Similar units are often located near one another so that phenomena of interest tend to cluster or exhibit similar patterns in space, making these phenomena spatially dependent. This clustering of outcomes (as well as any clustering in explanatory variables, including omitted variables that might lurk in the error term of statistical models) is no accident and should matter in how scholars understand and explain each outcome. That is, clustering is often a symptom of some underlying spatial process (e.g., diffusion), so scholars should take these underlying processes seriously in developing explanations of the related outcome of interest. Further, because the study of interdependent social phenomena is at the heart of social science, arguably all social science data are inherently spatial or—at the very least—spatial data are central to the social sciences (Darmofal 2015, 11–13). Nevertheless, the spatial dimensions of political data rarely receive explicit attention in how multimethod scholars design and conduct research.

Although both multimethod research and spatial analysis have gained popularity in recent years, the two methodological approaches have rarely entered into dialogue. This article identifies why this dialogue can be fruitful for both approaches—that is, integrating insights from both spatial analysis and the sequencing of quantitative and qualitative research can lead to better research designs and stronger conclusions than using either approach in isolation. Indeed, while there are other benefits from this integration, this article makes a core, two-pronged point: (1) without integrating insights from spatial analysis, multimethod designs can be self-defeating because one method may undermine the logic of another; and (2) without integrating insights from multimethod research, spatial statistics and econometrics can fall short by assuming rather than demonstrating both (a) the mechanisms underlying key spatial processes and (b) the proper unit or level of analysis. Because explicit strategies for integrating spatial econometrics and qualitative methods in multimethod research are discussed elsewhere, we refer readers to those works for specific strategies (e.g., Harbers and Ingram 2015; 2017a; 2017b) and focus instead on the more general motivations for combining spatial analysis with case-study research.

The discussion is organized in three sections that address three sets of reasons why multimethod scholars should explore the spatial dimensions of their data. To limit the scope, we focus on research designs in which large-N statistical analysis is combined with small-N qualitative analysis—an approach that recently has become particularly popular (Seawright 2016b; Seawright and Gerring 2008). The first section describes why the failure to consider spatial dependence may undermine inference in multimethod research. If spatial dependence exists but remains unaccounted for, multimethod designs are on shaky ground; scholars run the risk of missing part(s) of the causal process, which then undercuts the strength of their causal claims. The second section highlights what multimethod scholars can bring to spatial analysis and how multimethod research can be leveraged to better understand spatial processes. As the world becomes more connected and interdependent, effective methods to study dependence, diffusion, and spillovers—as well as corollary processes such as displacement and insulation—must become standard fare within social science toolkits. The third section highlights how incorporating the spatial nature of data can enrich multimethod research by providing a new set of geographic tools to analyze data. To illustrate this point, we show how basic, exploratory, univariate diagnostics of spatial autocorrelation can help scholars to (1) understand the boundedness of phenomena, and (2) select interesting cases for further analysis.

SPATIAL DEPENDENCE AS A THREAT TO INFERENCE IN MULTIMETHOD RESEARCH

Most comparative research designs build on the premise that the causal process unfolds within clearly delimited units and that these units are independent of one another. The idea that political boundaries delimit the core processes of politics seems so taken for granted and natural to many comparative scholars that it is rarely questioned, underpinning all of the classic works on comparative case-study design (e.g. Lijphart 1975; Przeworski and Teune 1970). These designs begin with the assumption that the causes of an outcome are contained within the unit in which it is observed, without explicit attention to characteristics of surrounding units. Whether cases are selected for in-depth analysis on the basis of characteristics of

the dependent variable (Y), the independent variables (X), or a specific configuration of the X-Y relationship, the independence of units generally is taken as so self-evident that this assumption is rarely explicitly stated, much less empirically probed. Multimethod scholars tend to follow in this tradition. For instance, in research in which case selection is informed by results of a prior regression analysis (Lieberman 2005; Seawright and Gerring 2008), units generally are assumed to meet the regression assumption of being "independent and identically distributed" (i.i.d.) and cases for in-depth analysis is an example of what some scholars might study as diffusion or convergence.

Although conceiving of units as principally independent or interdependent implies a different conceptualization of the causal process, most social science outcomes are likely due to a combination of unit-specific (aspatial) and contextual (spatial) variables (Cho 2003). Therefore, rather than conceiving of these two approaches as competing and mutually exclusive, for many research questions it is more productive to view them as complementary (Harbers 2017). The central

Spatial analysis begins from the opposite assumption: "[r]ather than considering N observations as independent pieces of information, they are conceptualized as a single realization of a process" (Anselin and Bera 1998, 252).

are studied individually and in isolation. Even small-N comparative research designs treat individual units as independent of other units.

Spatial analysis begins from the opposite assumption: "[r]ather than considering N observations as independent pieces of information, they are conceptualized as a single realization of a process" (Anselin and Bera 1998, 252).[1] Units—especially contiguous or nearby units—are assumed to be connected and a researcher is expected to know the nature of these interactions ex ante. Specifically, conducting spatial analysis requires the specification of an *n* by *n* weights matrix (W) that indicates whether a connection exists for all pairs of observations. In principle, the elements of this matrix can be specified on the basis of different criteria, ranging from strictly geographic conceptions (e.g., contiguity and Euclidean distance) to more social forms of connectedness (e.g., travel time, communication flows, and road networks) (Beck, Gleditsch, and Beardsley 2006). However, regardless of the form of connection, the core aim of spatial analysis is to uncover the type, magnitude, and reasons for dependence in the data.

Spatial analysts distinguish between two main types of spatial dependence: attributional and interactive. Attributional dependence refers to scenarios in which the similar within-unit characteristics of neighboring units make an outcome more or less likely in both (Darmofal 2015, 4); however, the shared attributes are not sufficiently understood to model them in a regression. That is, these attributes would not be identified, much less measured, and so would remain within the error term of a conventional regression. This type of dependence is associated with spatially correlated regression errors. Interactive spatial dependence occurs when geographic connections or interactions make an outcome more or less likely across connected units (Baller et al. 2001). This second process—similar to common notions of diffusion—is captured by a spatial lag of the dependent variable.[2] Depending on the analysis, either the error or the lag process may be more prominent. The broader point is that connected units interact with one another so that, over time, we are likely to see similar values emerge of not only an outcome of interest but also of other attributes as units influence one another. This spread of similar values

point is that in many social science analyses—especially in those in which relevant units are political jurisdictions—unit independence should not be assumed but instead explored empirically.

If spatial dependence exists, multimethod designs that ignore this dependence are on shaky ground. They do not apprehend the full causal process, which can threaten causal inference. In multimethod designs with a case-study component, the selection of cases for in-depth analysis assumes that researchers have an adequate understanding of the causal properties of a case (i.e., whether it is typical, deviant, or influential). The regression generally is used to summarize the data, and the case study provides the additional information necessary to go beyond description to make robust causal claims (Seawright 2016a, 45). However, if units are spatially dependent, and this dependence of the data is not modeled adequately, a regression is misspecified. At least, the misspecification limits its usefulness for case selection (Rohlfing 2008; Seawright 2016b, 500). Depending on the nature of spatial dependence, model parameters are biased and/or inefficient, even for the unit-specific, aspatial predictors in the model (Darmofal 2015, 32–3).

During the case-study phase, the investigated unit can tell only part of the causal story. If the quantitative phase detects spatial dependence, researchers cannot simply proceed by selecting units from the broader dataset for in-depth analysis. The existence of spatial dependence suggests that the causal process does not map neatly onto the considered units, which undercuts the benefits of an in-depth analysis of them. Ultimately, assuming unit independence can create blind spots that diminish the benefits of multimethod integration. In the worst case, case studies may reveal interdependence, which then would render the large-N model unconvincing and call into question the case-selection rationale, completely disconnecting the large-N phase from the small-N phase.

SPATIAL DEPENDENCE AS A TOPIC OF SUBSTANTIVE INTEREST

Whereas the previous section offered a cautionary tale about ignoring spatial dependence, this section highlights what

multimethod research can offer spatial analysis. Although it may be tempting to think of spatial dependence as exceptional in comparative research, there are sound reasons to believe that some level of spatial dependence is the rule whenever scholars work with aggregate data for political units. As noted previously, most outcomes in social science are clustered in space. In addition, the world today is arguably more interconnected than at any point in the past. Thus, the question of whether outcomes are influenced by events elsewhere deserves careful consideration. The increase in communication and population flows coincides with renewed interest in processes of

MOVING FORWARD: SPATIAL TOOLS FOR MULTIMETHOD RESEARCH

Finally, incorporating the spatial nature of social science data provides multimethod researchers with access to a new set of geographic tools that are useful even when a researcher's primary interest is not in the spatial process. To illustrate this broader point, this section highlights how a simple univariate examination of clustering in the outcome of interest can generate valuable insights. We outline two exploratory spatial statistics: the global and the local Moran's I statistics. The latter is a dominant form of a broader class of Local Indicators

The increase in communication and population flows coincides with renewed interest in processes of diffusion—and related phenomena such as transfer, learning, and contagion—within research agendas that seek to understand the spread of policies, violence, norms, and much else.

diffusion—and related phenomena such as transfer, learning, and contagion—within research agendas that seek to understand the spread of policies, violence, norms, and much else. Multimethod designs are uniquely placed to shed light on these phenomena and to answer questions about why similar units are located near one another, why a phenomenon of interest spreads to some units but not to others, and what role political borders and jurisdictions play in this process.

More specifically, spatial econometrics generally requires scholars to make strong ex ante assumptions about both the appropriate unit of analysis and the nature of dependence. That is, spatial analysts generally select units and specify the weights matrix (W)—which indicates whether and how intensely all pairs of units interact—at the beginning of the research process, rarely revisiting these specifications at later stages of the research cycle. Among other things, multimethod research provides important opportunities to test and, if necessary, update such regression assumptions (Seawright 2016a).

In the case-study phase, researchers who encounter interactive spatial dependence may seek to uncover "vectors of transmission"—that is, causal mechanisms or pathways that underlie detected spatial interactions (Baller et al. 2001). For instance, by process-tracing how an outcome spreads to other units, researchers can refine a weights matrix initially specified on the basis of contiguity and replace it with a more precise understanding of interactions, such as distance, travel time, and newspaper circulation.[3] By enabling scholars to revisit previous choices and to improve the econometric model on the basis of insights gleaned during the qualitative phase, multimethod designs provide scholars with additional leverage for enhancing the validity of their conclusions. In short, whereas spatial analysis improves qualitative research by focusing on relevant research questions and strengthening inferences related to them, qualitative research can strengthen spatial analysis by examining core assumptions about units of analysis and the nature of interdependence that generally are unexamined.

of Spatial Association (LISA) (Anselin 1995)—indeed, some authors use "local Moran" and "LISA" interchangeably—and can be used to (1) examine the boundedness of phenomena, and (2) select cases for in-depth analysis.

The global Moran's I statistic summarizes the overall degree of autocorrelation in the data and indicates whether the relationship between connected units is positive (i.e., connected units exhibit similar values) or negative (i.e., connected units exhibit dissimilar values). In addition, a local version of Moran's I, which is also called a LISA statistic, can identify local patterns of dependence and clustering.[4] A LISA map displays statistically significant clustering patterns including high-high clusters (in which units with high values are connected to other units with high values) and low-low clusters (in which units with low values are neighbored by units with low values). For these reasons, this type of map is among the most useful ways to explore spatial data. We draw on two examples to illustrate the value of the LISA statistic for multimethod research. The map shown in figure 1 encompasses all 2,455 municipalities in Mexico and graphs the clustering of an index that captures political participation in municipal elections.[5] The map shown in figure 2 encompasses all 94 subnational units (i.e., the first subnational administrative level) across seven Central American countries, graphing the clustering of homicide rates in the region.[6]

In case-study research, a unit is conceptualized as "a spatially bounded phenomenon—e.g., a nation-state, revolution, political party, election, or person—observed at a single point in time or over some delimited period of time" (Gerring 2004, 342). Although comparativists tend not to be concerned about the spatial boundedness of phenomena, the importance of identifying the appropriate scale for analysis is widely recognized. Gerring (2007, 19–20) addressed the issue explicitly, as follows:

> Note that the spatial boundaries of a case are often more apparent than its temporal boundaries…. Occasionally, the temporal boundaries of a case are more obvious than its spatial boundaries. This is true when the phenomena

under study are eventful but the unit undergoing the event is amorphous. For example, if one is studying terrorist attacks, it may not be clear how the spatial unit of analysis should be understood, but the events themselves may be well bounded.

The spatial boundaries of phenomena deserve close scrutiny not only when a scholar is studying singular events but also whenever phenomena are studied at some level of aggregation. More often than not, multiple levels of aggregation

served by using countries as their units. Homicidal violence, in this instance, is not bounded or delimited by country borders. In contrast, voter participation rates in Mexico (see figure 1) follow state borders more closely, even though the index captures participation in municipal rather than in state or federal elections. Ultimately, examining clusters allows researchers to not only reflect on the boundedness of phenomena but also to identify puzzles that inform new research questions, such as why some outcomes map onto political boundaries whereas others straddle the borders of multiple jurisdictions.

Mapping clusters and investigating the boundedness of phenomena provide opportunities to broaden the range of possible cases for in-depth analysis beyond individual political units.

are possible. Although the causes of a particular homicide are likely to be the purview of the police, the question of why violent crime is more prevalent in some political units (i.e., neighborhoods, municipalities, states, or countries) than in others is of interest to political scientists and other social scientists, and to policy makers representing those units.

Mapping LISA clusters of aggregate data allows researchers to reflect on the appropriate unit or scale of analysis for case-study research. Figure 2 reveals a cluster of high homicide rates encompassing parts of Guatemala, Honduras, and El Salvador. This area is known as the Northern Triangle, where the high incidence of violence already has forced many citizens to flee. Because the cluster extends across the three countries but encompasses only parts of each country, researchers seeking to identify the causes of this violence may be poorly

Mapping clusters and investigating the boundedness of phenomena provide opportunities to broaden the range of possible cases for in-depth analysis beyond individual political units.

Notably, a thorny and persistent challenge in spatial analysis is the modifiable areal unit problem (MAUP) (Openshaw and Taylor 1979; Wong 1996). This problem refers to the fact that estimates of spatial patterns may change depending on the number and type of units chosen. There is no technical solution to this problem; therefore, the best current advice for spatial analysts is to be mindful of this issue and to select units that are consistent with their theory and research interests (Darmofal 2015, 26–7). By identifying the proper unit or level of analysis, these quantitative tools also help address MAUP. Moreover, the in-depth, qualitative research conducted within selected cases can also help develop a better understanding of the proper unit or level of analysis, further resolving challenges associated with MAUP.[7]

Inspecting data spatially also can be useful for puzzle-driven case selection, especially when clusters appear in unexpected locations. In light of the extensive literature on violence in the Northern Triangle, for instance, the low-low cluster in Guatemala warrants closer scrutiny. Broadly, the existence of clusters raises questions about the origins of spatial dependence, and multimethod research can contribute to a better understanding of observed

Figure 1

LISA Cluster Map of Voter Participation Rates in Mexico, 2010
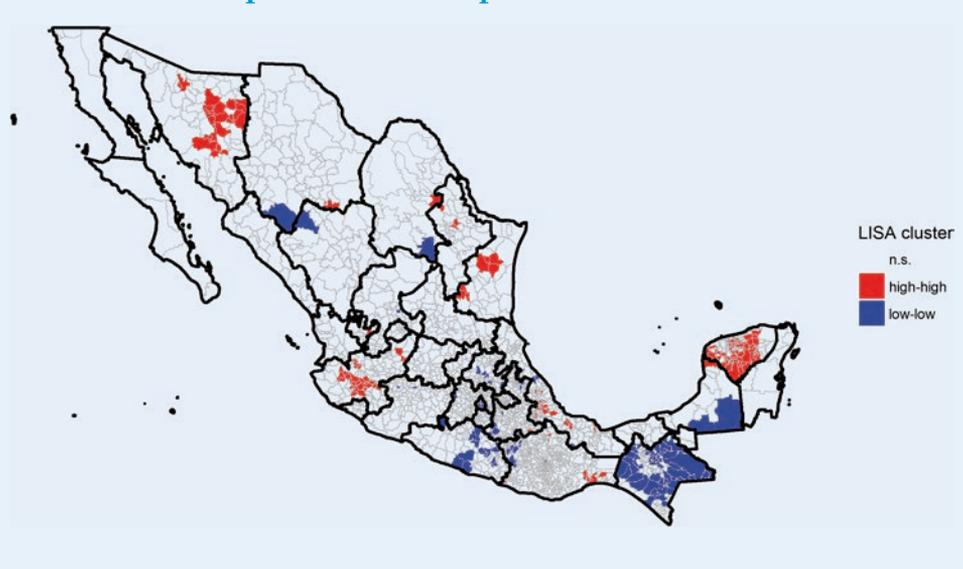
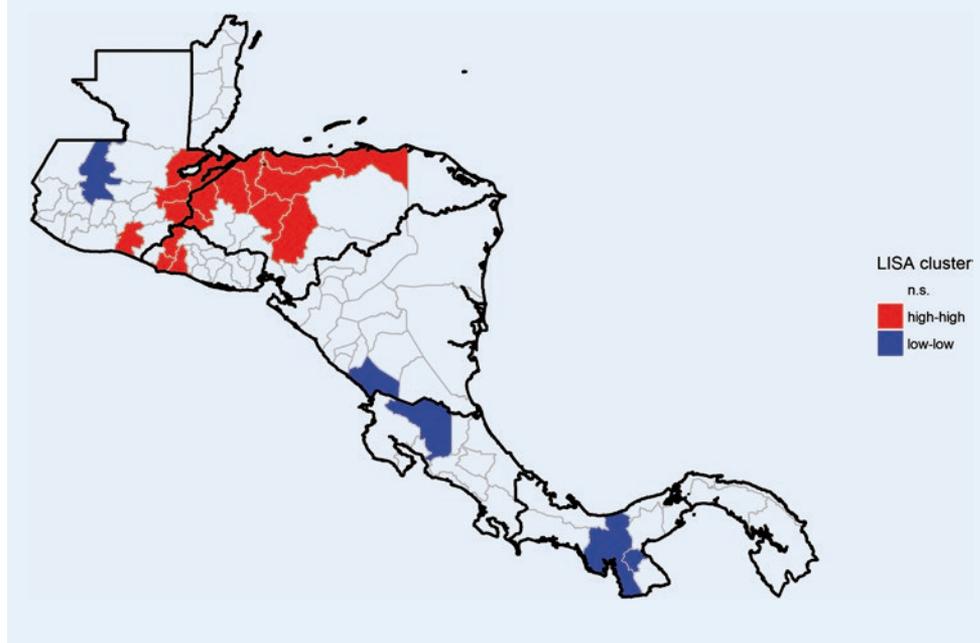

LISA cluster
n.s.
high-high
low-low

Figure 2

LISA Cluster Map of Homicide Rates in Central America, 2010

allows relationships to vary across space, there may be little that the mathematical model can illuminate regarding the spatially varying coefficients produced by GWR. A logical next question is to identify which other third variable might be exerting a conditioning or moderating effect to produce this local variation. However, qualitative work and case knowledge can provide explanations for this variation (Harbers and Ingram 2017a; 2017b).

The main objective of this article is to highlight why dialogue between spatial analysis and multimethod research is promising. Going forward, we hope that interaction—rather than independence— will inform the relationship between these two research traditions.

spatial patterns. Analysts can choose a set of units from each LISA cluster category—high-high, low-low, and nonsignificant areas—which then could be combined in a small-N research design to explore diffusion, barriers to diffusion, omitted variables, and other underlying dynamics across the full range of clustering patterns. Notably, the notion of a "case" must be rethought if there is evidence of interdependence. That is, it makes little sense to examine cross-unit diffusion or interactions of any kind by focusing on a single unit in isolation. Rather, a case should consist of at least two connected units, thereby enabling researchers to examine the nature of the connection and cross-unit interaction.

Leveraging clusters for case selection and using them to examine the boundedness of phenomena are only two examples of how even a simple exploratory tool from spatial analysis can strengthen multimethod designs. There are many other ways in which spatial analysis can enrich the toolbox of multimethod scholars.

As another example, spatial analysts take seriously the possibility of spatial heterogeneity—that is, the relationship between an explanatory variable (X) and the outcome of interest (Y) may not be uniform across space. Conventionally, a regression coefficient captures the extent to which X and Y move together, and researchers report a single coefficient for all observations. Spatial heterogeneity implies that a single regression coefficient is inadequate; rather, local, unit-specific coefficients may be necessary to capture variation in the magnitude, direction, and statistical significance of the relationship between X and Y (e.g., Brunsdon, Fotheringham, and Charlton 1996).[8]

Qualitative research can provide considerable insight into the sources of this heterogeneity. For example, when scholars employ geographically weighted regression (GWR), which

NOTES

1. In this regard, the focus on dependence in spatial analysis is conceptually and analytically similar to other approaches with dependent data structures (e.g., time-series analysis and network analysis).

2. Empirically, both types of dependence may be present at the same time. The clustering of democracy, for instance, is due to the fact that wealthy countries and those with favorable institutional preconditions are located in the same world regions as well as a neighborhood effect because democracies nearby increase the likelihood of democracy locally (Brinks and Coppedge 2006). Terminology about the nature of dependence varies among scholars, especially across different disciplines (e.g., "substantive dependence" in Baller et al. 2001).

3. For a more in-depth discussion of how multimethod research may be leveraged to understand the spatial process, see Harbers and Ingram 2017a and 2017b.

4. Local indicators of spatial association in addition to the local Moran are also available, including local Gamma and local Geary (Anselin 1995).

5. The index is constructed by dividing the number of actual voters by the number of registered voters in the two previous municipal elections (i.e., *votos emitidos/lista nominal*) (Flamand, Martinez Pellegrini, and Camacho 2007).

6. Originally reported in Ingram and Curtis 2015.

7. We thank an anonymous reviewer for encouraging us to discuss the added value of qualitative research in addressing MAUP.

8. See Ingram and Marchesini da Costa 2017 for an application of this method to the study of violence in Brazil.

REFERENCES

Anselin, Luc. 1995. "Local Indicators of Spatial Association—LISA." *Geographical Analysis* 27 (2): 93–115.

Anselin, Luc, and Anil K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of applied economic statistics*, eds. Aman Ullah, David Giles, 237–289. New York: Marcel Dekker.

Baller, Robert D., Luc Anselin, Steven F. Messner, Glenn Deane, and Darnell F. Hawkins. 2001. "Structural Covariates of US County Homicide Rates: Incorporating Spatial Effects." *Criminology* 39 (3): 561–88.

Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006. "Space Is More Than Geography: Using Spatial Econometrics in the Study of Political Economy." *International Studies Quarterly* 50 (1): 27–44.

Brinks, Daniel, and Michael Coppedge. 2006. "Diffusion Is No Illusion: Neighbor Emulation in the Third Wave of Democracy." *Comparative Political Studies* 39 (4): 463–89.

Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton. 1996. "Geographically weighted regression: A method for exploring spatial nonstationarity." *Geographical Analysis* 28 (4): 281–298

Cho, Wendy K. Tam. 2003. "Contagion Effects and Ethnic Contribution Networks." *American Journal of Political Science* 47 (2): 368–87.

Darmofal, David. 2015. *Spatial Analysis for the Social Sciences*. New York: Cambridge University Press.

Flamand, Laura, Sarah Martinez Pellegrini, and Ofelia Camacho. 2007. "Metodología de Cálculo: Índice de Deasarrollo Municipal Básico (IDMb)." *El Colegio de la Frontera Norte*. Available at www.inafed.gob.mx/work/siha_2015/4/MetodologiaIDMb-Feb07.pdf.

Gerring, John. 2004. "What Is a Case Study and What Is It Good For?" *American Political Science Review* 98 (2): 341–54.

——. 2007. *Case Study Research: Principles and Practices*. New York: Cambridge University Press.

Harbers, Imke. 2017. "Spatial Effects and Party Nationalization: The Geography of Partisan Support in Mexico." *Electoral Studies*. Available at http://doi.org/10.1016/j.electstud.2016.11.016.

Harbers, Imke, and Matthew C. Ingram. 2015. "Politics in Space: Methodological Considerations for Taking Space Seriously in Subnational Comparative Research." Paper presented at the American Political Science Association Annual Meeting.

——. 2017a. "Spatial Tools for Case Selection: A Menu of Mixed-Methods Research Designs." Paper presented at the Southern Political Science Association Annual Meeting.

——. 2017b. "Geo-Nested Analysis: Mixed-Methods Research with Spatially Dependent Data." *Political Analysis*. DOI: 10.1017/pan_2017.4

Ingram, Matthew C., and Karise M. Curtis. 2015. "Violence in Central America: A Spatial View of the Region, Northern Triangle, and El Salvador." In *Crime and Violence in Central America's Northern Triangle: How U.S. Policy Responses are Helping, Hurting, and Can Be Improved*, ed. Eric Olson, 245–83. Washington, DC: Woodrow Wilson Center.

Ingram, Matthew C., and Marcelo Marchesini da Costa. 2017. "A Spatial Analysis of Homicide Across Brazil's Municipalities." *Homicide Studies* 21 (2): 87–110.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52.

Lijphart, Arend. 1975. "The Comparable Cases Strategy in Comparative Research." *Comparative Political Studies* 8 (2): 158–77.

Openshaw, Stan, and P. J. Taylor. 1979. "A Million or So Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem." In *Statistical Applications in the Spatial Sciences*, ed. N. Wigley, 127–44. London: Pion Limited.

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley & Sons.

Rohlfing, Ingo. 2008. "What You See and What You Get: Pitfalls and Principles of Nested Analysis in Comparative Research." *Comparative Political Studies* 41 (11): 1492–514.

Seawright, Jason. 2016a. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge, UK: Cambridge University Press.

——. 2016b. "The Case for Selecting Cases That Are Deviant or Extreme on the Independent Variable." *Sociological Methods & Research* 45 (3): 493–525.

Seawright, Jason, and John Gerring. 2008. "Case-Selection Techniques in Case-Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61 (2): 294–308.

Wong, David. 1996. "Aggregation Effects in Geo-Referenced Data." In *Practical Handbook of Spatial Statistics*, ed. Sandra Arlinghaus, 83–106. Boca Raton, FL: CRC Press.