Can Reproducible and replicable research facilitate causal explanation in Geography?

Peter Kedron
Arizona State University
5592 Coor Hall
Tempe, AZ, USA
Peter.Kedron@asu.edu

Abstract

Studies examining the reproducibility of geographic research suggest that it may be difficult to reproduce many published findings. While a small, but growing, body of work is defining reproducible practices within GIScience, it has engaged less with the replication of geographic research and the explanation of geographic phenomena. This paper addresses this gap in the literature by examining the benefits and limits of reproducible and replicable research in the explanation of geographic phenomena. Specifically, this paper focuses on how two fundamental properties of geographic phenomena, spatial heterogeneity and spatial dependence, affect the development of causal explanations and discusses how reproducible practices might improve this process.

Keywords: Reproducibility, Replicability, Causal Explanation, Spatial Heterogeneity, Spatial Dependence, GIScience

1 Replication in Geography

The ability to reproduce and replicate prior findings is essential to self-correction in scientific inquiry. Although the precise definition of reproduction and replication vary by discipline (Barba 2018; Plesser 2018), the American Statistical Association differentiates the two terms by defining reproducibility as use the original data and methods to recreate the results of a study, and replicability as independently repeating a study with the original data, but generally using the same methods and finding corroborating results (Broman et al. 2017). Given the role reproducibility and replicability play in the practice of science, the inability to reproduce and replicate prior results in a range of fields (Ioannidis et al. 2005; Baker 2016; Camerer et al. 2016; Ioannidis et al. 2017) has expanded research into the performance, communication, and evaluation of research itself - meta-research (Ioannidis 2018).

While geography has a long tradition of examining the research process and the formation of theory (see Hartshorne 1939; 1955; Schaefer 1953; Harvey 1969; Martin 2005), the reproducibility and replicability of geographic research has yet to be broadly discussed in this present context (Nust et al. 2018). To date, a small number of studies analysing the reproducibility of geographic research have (i) demonstrated that it is currently challenging to reproduce published work (Konkol et al. 2019) and (ii) identified and encouraged the adoption of research practices that facilitate reproduction of geographic findings (Brundsdon and Singleton 2015; Brunsdon 2016; Arribas-Bel and Reades 2018; Granell et al. 2018; Nust et al. 2018). Both activities are necessary steps toward improving the transparency, portability, comparability, and impact of geographic research. However, research has primarily emphasized reproduction over replication, and only indirectly addressed how the unique characteristics of geographic phenomena may influence the effectiveness of selected practices. To amplify the impact of reproducible practices, those practices should be placed in disciplinary context and connected to a primary goal of geographic inquiry – the formation of theory and the explanation of geographic phenomena.

This paper begins to address these needs by examining the benefits and limits of reproducible and replicable research in the explanation of geographic phenomena. Specifically, this paper focuses on how two fundamental properties of geographic phenomena, spatial heterogeneity and spatial dependence, affect the development and testing of causal explanations using reproducible practices. While a comprehensive treatment of these issues is beyond the scope of a single work, this paper aims to highlight factors that complicate causal explanation while also raise awareness of reproducible practices.

2 Causal Inference

Harvey (1969) identifies six forms of analysis used to develop and formalize the explanation of geographic phenomena: cognitive, morphometric, causal, temporal, ecological, and systems. This paper focuses on causal analysis because this mode of explanation may be flexibly applied to a wide range of situations. For example, once an understanding of cause and effect has been established between two variables, that knowledge can be used to analyse individual outcomes, examine dynamic systems, and build theory. However, causal logic can only be applied in empirical research when (i) the variables participating in and confounding a causal relationship are identified, and (ii) the sections of reality in which that relationship holds have been established.

The primary barrier to identifying either the boundaries or variables of a causal relationship is that causal effects cannot be directly observed when defined at a selected unit of analysis (Berk 2005). For example, an individual can never be simultaneously assigned to both receive and not receive medical treatment. Ideally, to determine the causal effect of an intervention the individual outcome from both the observed treatment and the unobserved counterfactual non-treatment would be known. In the experimental sciences, the common solution is to move to the group level and use a controlled experiment and random assignment to balance individuals receiving interventions and not receiving interventions on all confounders. This approach allows for estimation of an average response to the intervention, which can be used as an estimate of the causal effect for a type of individual, while not the individuals themselves. By adjusting and repeating the experiment, the boundaries of the causal relationship can be identified and the variables involved in its function can be determined (Berk 2005).

Using the above procedure to identify a causal relationship when studying geographic phenomena is difficult because it assumes (i) that intervention is possible and (ii) that, if an intervention occurs, the condition to which an individual is assigned has no impact on the response of another individual (the stable unit treatment value assumption). In geography, direct interventions and randomized assignment are rare. Without direct intervention the size of the treatment can be difficult to identify, and without random assignment the influence of confounding factors can be difficult to determine. Instead, causal understanding is built through a combination of abductive, inductive, and deductive inference and the linking of the resulting theoretical structures to real world situations (Harvey 1969; Goodchild 2004a; 2004b). Even in instances where interventions and randomization are possible, geographic phenomena tend to exhibit two general properties, spatial heterogeneity and spatial dependence, which make it likely that the response of any individual unit will be at least partially affected by those surrounding it. The following section discusses these two properties and their relationship to causal inference and the reproducibility and replicability of geographical analysis in detail.

3 Properties of Spatial Data

Anselin (1989) identifies two general properties of spatial data which complicate the analysis of geographic phenomena: spatial heterogeneity and spatial dependence.

3.1 Spatial Heterogeneity

Spatial heterogeneity refers to the uneven distribution of a variable across space, and is generally considered a result of the structural differences that exist between locations (Anselin 2010). The existence of spatial heterogeneity has at least two important consequences for the identification of causal effects. First, by its definition spatial heterogeneity indicates that causal mechanisms and confounding variables will vary in space, and by extension that the concept of an average place does not exist (Goodchild 2004a). Second, because causal mechanisms differ with location, the results of any analysis must therefore depend on the geographic boundaries of that analysis (Goodchild 2004b). Taken together, the two consequences of spatial heterogeneity suggest that identifying the variables involved in a causal relationship and the geographic contexts in which that relationship holds will not yield a fixed result that is universally applicable across space.

Expecting causal mechanisms to potentially vary across space has further implications for the role of the reproduction and replication of spatial analysis in the explanation of geographic phenomena. First, spatial heterogeneity suggests that the baseline assumption of a replicated analysis should be one of non-replication, or replication with a different set of confounding influences. For example, using the same methods to examine the causes of poverty in large metropolitan area and a small rural town in the United States is likely to omit important variables, or even completely misspecify the causal mechanism in at least one location. Even in the case where an intervention with random assignment to the treatment and controls is the basis of the replication, spatial heterogeneity could still produce different results. Second, moving to the regional level and treating locations, rather than individuals, as the unit of analysis does not solve the problem because the identified average causal effect of an intervention will have limited application in heterogeneous regions.

Nonetheless, reproduction and replication do have an important role to play in establishing of causal effects in geographic research. In the presence of spatial heterogeneity it is important to trace the geographic boundaries of causal effects. This is a primary motivation of many forms of place-based analysis such as geographically weighted regression (Fotheringham et al. 2002), spatial applications of multilevel modeling (Rabash et al. 2014), Local Indicators of Spatial Association (Anselin 1995), and spatial scan statistics (Graz et al. 2009). Developing and adhering to standards of reproducible spatial analysis when applying these methods will help characterize the boundaries of potential causal relationships and the variables they involve. While cataloguing variables and boundaries during replications cannot definitively establish a causal effect, it can set the stage for targeted randomized interventions in selected locations that could then lead to the development of geographic theories capable of covering those places.

3.2 Spatial Dependence

Spatial dependence, also known as Tobler's First Law, is the "the propensity for nearby locations to influence each other and to possess similar attributes" (Goodchild 1992, 33). In other words, things that are close together tend to be more related than things that are far apart. Spatial dependence complicates the identification of causal effects because influential relationships among individual observations make parameter estimates unreliable and, in the presence of a randomized intervention, violate the stable unit treatment value assumption.

Even in the presence of a spatially homogenous causal mechanism, with a known set of confounding variables, and a randomized treatment, spatial dependence obscures causal effects. The problem created by spatial dependence is that, for any individual, the potential response to the treatment or control can vary as a function of the treatment assignment of other individuals. For example, if an educational intervention is introduced to a neighbourhood to reduce poverty, because individuals assigned to the treatment or control groups are likely to interact, the skills gained by those in the program could be passed to those assigned to the control group, or amplified by interactions with others in the treatment group. How skills are shared depends on the spatially dependent pattern of interactions within the neighbourhood. Critically, simple random assignment of the intervention does not correct this problem. In the presence of spatial dependence, any pattern of assignment can change the observed response because each intervention will place skills at different points in the interaction network, leading to different sharing and response patterns. Definition of the causal effect is then dependent on the patterns compared.

More often in geographic research, intervention is not possible, and researchers are forced to rely on assignment to 'treatments' in natural experiments that are themselves likely influenced by spatial dependencies. If spatial dependencies can be evaluated and adjusted for, analysing the response of individual locations matched on confounding variables can give an estimate of causal effect. However, as in the case of a direct intervention, our inability to completely model the nature of spatial interactions is likely to produce only one of many possible estimates. Current, spatial statistical methods facilitate diagnoses of the magnitude and scale of spatial dependence (e.g., Moran's I, Ripley's K), and adjust for its influence on parameter estimation (e.g., spatial regression). However, researchers generally do not use these methods to create detailed characterizations of the interactions influencing confounding or causal variables. For example, it is only recently that relationships between spatial and non-spatial dependencies have been explored in regression frameworks applied in economic geography (Marrocu et al. 2013, Chiara et al. 2016).

Improving the reproducibility and replicability of geographic research can help address the effects of spatial dependence on the development of causal explanations. Reproducible research that clearly outlines how spatial dependencies are conceptualized and adjusted for will help others evaluate the plausibility of an estimated causal effect. Similarly, clearly outlining how spatial dependencies are modelled as processes of diffusion, mixing, interaction, or dispersal (Haining 2003) will facilitate the replication of methods in new studies and identification of the boundaries of casual relationships. Transparency in the process form used to capture dependencies is particularly important for replication because exact replication of a spatial weights matrix or density estimator (the most common means of capturing spatial dependence in spatial statistical analysis) is of little use in geographic research. As with spatial heterogeneity, repeated evaluation and tracking of the processes that produce spatial dependence and the situations in which they appear to have an effect can pave the way for targeted direct interventions where adjustments can be made for the influence of one location on another.

4 Conclusion

Producing causal explanations depends on identifying the variables involved in a causal relationship and defining the sections of reality in which that relationship holds. The spatially heterogeneous and spatially dependent nature of geographic phenomena complicate both of these tasks. The existence of spatial heterogeneity implies that causal effects and confounders should not be expected to be stable across space, while the presence of spatial dependencies makes it difficult to separate a causal effect from the spatial interactions that might confound its magnitude and spatial pattern.

These factors have several implications for the reproducibility and replicability of geographic research and the role those activities can play in the explanation of geographic phenomena. First, in the presence of spatial heterogeneity and dependence there is no reason to expect that replications of geographic studies will necessarily universally produce corroborating results. Even replications of randomized interventions in the same geographic context could be expected to produce different results if spatial dependencies have changed. Second, reproduction and replication nonetheless have important roles to play in defining the boundaries within which that causal relationships hold and identifying the variables involved those relationships. As in the experimental sciences, thoughtful examination and precise attempts at replication are only possible if the details of prior studies are fully shared and accessible. With the full details of prior research designs available, a body of observational replication attempts could be created that collectively outline the geographic boundaries and confounding variables of a causal effect. Finally, that focused body of knowledge could then set the stage for targeted interventions with appropriate adjustments for spatial heterogeneities and dependencies. These type of direct interventions remain rare in geographic research, but have the potential to contribute greatly to our understanding of causal effects.

References

Anselin, L. (1989) What is special about spatial data? Alternative perspectives on spatial data analysis, Technical Report 89-4. Santa Barbara, CA: National Center for Geographic Information and Analysis.

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

Anselin, L. (2010) Thirty years of spatial econometrics. *Papers in Regional Science* 89(1): 3–25.

Arribas-Bel, D., Reades, J. (2018) Geography and computers: past, present, and future. *Geography Compass 12*(10): e12403.

Baker, M. (2016) 1500 scientists lift the lid on reproducibility. *Nature* 533(7604): 452-454.

Barba, L. (2018). Terminologies for reproducible research. *arXiv preprint arXiv:1802.03311*

Berk, R. (2005). Randomized experiments as the bronze standard. *Journal of Experimental Criminology* 1(4): 417-433.

Broman, K., Cetinkaya-Rundel, A., Nussbaum, C., Paciorek, R., Peng, D. Turek, Wickham, H. (2017) Recommendations to funding agencies for supporting reproducible research. American Statistical Association, http://www.amstat.org/asa/files/pdfs/POL-ReproducibleResearchRecommendations.pdf.

Brunsdon C., Singleton A (2015) Reproducible research: Concepts, techniques and issues. In: Brunsdon C., Singleton A (eds), Geocomputation: A Practical Primer. Sage, London, pp. 254–64.

Brunsdon, C. (2016). Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography* 40(5): 687-696.

Camerer, C., Dreber, A., Forsell, E., Ho, T., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science* 351(6280): 1433-1436.

Chiara, M., Marrocu, E., Paci, R. (2016) The concurrent impact of cultural, political, and spatial distances on international mergers and acquisitions. *The World Economy* 39(6): 824-852.

Fotheringham, S., Brunsdon, C., Charlton, M. (2002) *Geographically weighted regression: The analysis of spatially varying relationships*. New York: Wiley.

Goodchild, M. (1992) Geographical information science. *International Journal of Geographic Information Systems* 6(1): 31-45.

Goodchild, M. (2004a) GIScience, geography, form, and process. *Annals of the Association of American Geographers* 94(4): 709-714.

Goodchild, M. (2004b) The Validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94(2): 300-303.

Granell, C., Nüst, D., Ostermann, F., Sileryte, R. (2018). Reproducible Research is like riding a bike (No. e27216v1). *PeerJ* Preprints.

Graz, J., Pozdnyakov, V., Wallenstein, S. (Eds) *Scan Statistics: Methods and Applications*. London, Springer.

Haining, R. (2003) Spatial Data Analysis: Theory and Practice. Cambridge, Cambridge University Press.

Hartshorne, R. (1939) The character of regional geography. *Annals of the Association of American Geograhers* 29(4): 436-456.

Hartshorne, R. (1955). "Exceptionalism in Geography" Re-Examined. *Annals of the Association of American Geographers* 45(3): 205-244

Harvey, D. (1969). *Explanation in Geography*. London, Edward Arnold.

Ioannidis, J. (2005) Why most published research findings are false. *PLOS Medicine* 2(8): e124.

Ioannidis, J., Stanley, T., Doucouliagos, H. (2017) The power of bias in economics research. *The Economic Journal* 127(605): F236-F265.

Ioannidis, J. (2018). Meta-research: why research on research matters. *PLOS Biology* 16(3): e2005468.

Marrocu, E., Paci, R., Usai, S. (2013). Proximity, networking and knowledge production in Europe: what lessons for innovation policy? *Technological Forecasting and Social Change* 80(8): 1484-1498.

Martin, G. (2005) *All Possible Worlds*. Oxford University Press, Oxford.

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., Cerutti, V. (2018). Reproducible research and GIScience: an evaluation using AGILE conference papers. *PeerJ*, *6*: e5072.

Plesser, H. (2018) Reproducibility vs. replicabilityL a brief history of a confused terminology. *Frontiers in Neuroinformatics* 11: 76. doi: 10.3389/fninf.2017.00076

Rabash, J., Browne, J., Healy, M., Cameron, B., Charlton, C. (2014). MLwiN Version 2.30. *Centre for Multilevel Modelling, University of Bristol*.

Schaefer, F. K. (1953). Exceptionalism in Geography: A Methodological Examination. *Annals of the Association of American Geographers* 43(3): 226-249.