Uncertainty in Causal Neighborhood Effects: A Multi-Agent Simulation Approach

Cécile de Bézenac
University of Leeds, UK
The Alan Turing Institute, London, UK

Abstract

Interaction between individuals within an environment can result in complex patterns that a statistical analysis is unable to disentangle. The resulting social structure may pose important challenges for the identification of causal relations between variables using only observational data. In particular, the estimation of contextual or neighborhood effects will depend on the spatial configuration under study and the morphology of the areas used to define them. The relevant interpretation of estimates is hence put into question. I suggest adopting a Agent Based Modeling (ABM) approach to study the uncertainty of neighborhood effect estimations within complex spatial systems. An Approximate Bayesian Computing algorithm is used to quantify the uncertainty on the underlying processes that may lead to such estimations. An ABM model of spatial segregation is implemented to illustrate this method.

2012 ACM Subject Classification Human-centered computing

Keywords and phrases Spatial causal inference, neighborhood effects, uncertainty, Agent Based Modeling, Pattern Oriented Modeling

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.26

Category Short Paper

1 Introduction

The endeavour to generate causal rather than associational claims requires identifying the sufficient conditions for real effects to be estimated. However, their identification can prove to be very challenging in the presence of social and spatial complexity. Namely, interaction between individuals within a geographic environment can result in spatial patterns that a statistical analysis is unable to disentangle. Causal studies that consider factors of a spatial nature however rarely question the nature of space or the different ways in which it may transform the causal analysis [10].

In this paper I consider a category of spatial exposures, the neighborhood effects: the causal effect on an outcome of living in a given area versus living elsewhere. The underlying questions that such exposures raise concern the implications of spatial assumptions on the causal nature of estimated effects. In particular, how does the way we assign neighborhoods to a delimited area impact our causal claims? This question interrogates the inherent uncertainty linked to space in worlds where individuals are in constant movement, in interaction with their peers and with their surroundings. I present some of the challenges linked to the estimation of area-level effects in observational studies. I propose an Agent Based modeling (ABM) approach to generate various forms of spatial complexity against which statistical models may be tested. The exploration of spatial configurations represents an interesting starting point to analyse the so-called neighborhood effects. In particular, ABM exploration methods such as Approximate Bayesian Computing (ABC) offer the means to evaluate the relevance of spatial estimands given complexity assumptions. An illustration of this approach is presented using a simple Schelling model of segregation.

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 26; pp. 26:1–26:6
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

2 Challenges of estimating and interpreting neighborhood effects

2.1 What are neighborhood effects?

In the literature, one commonly refers to neighborhood effects as the independent effect of a neighborhood on one or multiple outcome [8, 4] [1]. For instance, epidemiology studies may compare health outcomes in "poor" versus "rich" areas [17]. Behind this definition lies a number of strong assumptions: what spatial attributes (shape and scale) best describe the neighborhood given the research question? Through what mechanisms can this geometric object be thought to have an independent effect on individual outcome? Finally what are the conditions for these area-level effects to be interpreted causally?

Multiple causal pathways may generate a dependence between the spatial configuration and the individual outcome of interest [4]. The social structure can be thought to play an important role in the shaping of place and the creation of an area identity. Furthermore, social influence or contagion mechanisms may exacerbate or spread exposure effects within an area. Other forms of spatial processes may impact individuals locally such as pollution, crime or the presence of green spaces, etc. The combination of these spatial and social processes can generate important spatial heterogeneity that is typically approximated by variability between specific neighborhood attributes.

2.2 Challenges for estimation

Methods have been developed to estimate neighborhood effects and account for spatial heterogeneity [3]. Namely, multi-level regression models assume some spatial hierarchical structure and allow for heterogeneity by including both fixed and random area-level effects [8]. Still, the interpretation of results are exposed to serious challenges.

The identification of such effects is threatened for one by spatial confounding and at times, complete confounding. A selection process may render the comparison of observations in different areas impossible [1]. This process is further enhanced by social interaction and the non-independence of observations. In a causal analysis, this is referred to as interference or spill-over effects [13].

Finally, spatial and social phenomena can rarely be confined within fixed, arbitrary geographic borders. The choice of spatial areas can lead to the misspecification of neighborhoods that may not reflect any empirical reality. This geographic problem is known as the Modifiable Areal Unit Problem (MAUP) [6].

The previous challenges suggest that the primary danger with mapping the social onto fixed spatial boundaries does not pertain so much to the approximation as it does to the causal interpretation of these so-called neighborhood effects. While accounting for the latter has proven to be insightful for the study of health outcome, employment or education [16, 9], very little work has specifically considered the uncertainty introduced by spatial assumptions. The relation between the misspecification of spatial properties and the bias in results should be looked into. The potential of ABM to generate and analyse the uncertainty surrounding estimates is presented below.

[1] In this paper, we interchangeably use the terms area effects and neighborhood effects when considering small scale spatial configurations

3 A multi-simulation approach

In order to generate artificial complex systems from which granular data may be extracted, we adopt an Agent Based Model (ABM) approach. These models are particularly adapted to create emergent properties from the bottom up by allowing the modeler to build heterogeneous interacting agents while maintaining full control over the micro-level process [12]. The output of these ABM can be placed under the microscope of the same statistical models typically used in observational studies, of which: models that include spatial neighborhood effects.

Many validation methods have been developed to evaluate the performance of ABM for modelling real world systems. Some of these are able to integrate both empirical information and some level of uncertainty on the underlying process. One such method is Approximate Bayesian Computing that approaches a posterior distribution for the parameters of the ABM through typical MCMC algorithms [7]. The general idea is that given information on the system one is modeling, a distribution over the parameter space can be proposed to reflect the probable worlds that may have resulted in similar observations. Samples of the input space are drawn and either selected or rejected according to a proximity criteria, usually determined by an error threshold ϵ. This threshold represents a level of uncertainty in the output space: how precise is the information on the real system? A more detailed description of ABC can be found in [14].

I suggest using ABC, not to evaluate the ABM but to interrogate the relevance of spatial assumptions in analytical models. The questions that this framework should be able to answer are: How biased are the estimations of neighborhood effects when obtained from data generated by the ABM? What other complex systems and mechanisms may be considered as possible generators of these estimations, given the statistical confidence levels. The notion of *sufficiency* of the statistical approach is introduced here as its performance in distinguishing between multiple causal mechanisms can be analysed. The uncertainty considered here links the *equifinality* of complex systems [15] to the epistemic uncertainty of classical models [2]. In the following, I present a simple illustration of the use of this framework on a spatial segregation model, the Schelling model. This model was chosen as an illustration for it's simplicity and its focus on spatial interaction between agents.

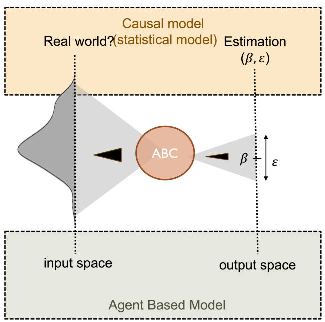


Figure 1 Schema of the ABM-ABC framework. Here β represents the objective/estimation and ϵ the threshold/standard error.

4 Illustration and results

4.1 The Schelling model

The Schelling model [11] can be described in the following way: on a square lattice a number of blue (exposed) and red pawns (non-exposed) occupy individual cells. These pieces will move to an empty cell if the percentage of a piece's same-color neighbors (Moore neighborhood) is lower than a predetermined threshold H (designating the homophily level). At each step, the elements on the board will be displaced according to their surrounding composition until they can no longer move or until all pieces are satisfied. The dynamics of this system tell a very interesting story as clear segregation patterns emerge without any higher order intervention. The relationship between the agent's attributes and their environment can be simply translated into a linear form for a given step in time.

$$Y_i(H, C_i, C_{N_i}) = H - C_i - \frac{1}{d_i} g(C_{N_i}) + \frac{2C_i}{d_i} g(C_{N_i}) \tag{1}$$

Where Y_i is the outcome for agent i, C_i is their color or the individual "exposure", $g(C_{N_i})$ is the aggregated prevalence of blue within i's ego-neighborhood and finally d_i is size of i's neighborhood. Translated into network terms, d_i is the degree of i in the regular graph drawn from the grid structure (for Moore neighborhood, $d_i \leq 8$). Note that there is clear violation of the no-interference assumption as agents outcome will depend on their own color and the color of peers. I consider the simple data scenario where the graph (or ego) neighborhood of pieces is not known but interaction is assumed constrained to fixed predetermined areas. Neighborhood exposure is then approximated by "blue" prevalence within an area. This specification of neighborhoods is analysed in light of the approximated posterior distribution for two of the models' key parameters: the homophily level (H) and the probability of agents belonging to the blue group (p). Some of the results obtained for different neighborhood scales are presented below.

4.2 Results

The 40x40 Schelling torus grids are cut up into 4, 25 and 100 different areas, respectively. An ABC is run for each choice of scale:

Samples were drawn from a uniform prior over the input space. A linear estimation of exposure and neighborhood effects are computed on a random simulation output for $p = 0.5$, $H = 0.5$. The estimation at step $t = 30$ is used as the objective for the ABC algorithm, the standard error of the estimation serves as the restrictive threshold for the rejection process. N=1000 particulars are tested against these criteria. Both the ABM model and the ABC algorithm were run using standard Python packages [5].

The estimation of color and neighborhood effects vary with time as the spatial patterns converges to a segregated state. It is quite clear that the choice of scale impacts the quality of the estimation. Larger scales blur the information on micro-level heterogeneity and social borders are hidden within fixed areas. The problem of total confounding may also arise for smaller scales as more and more selective neighborhoods appear. The results of the ABC show that the uncertainty introduced by the spatial approximation does interact with model parameters in a uniform way. While the true homophily level is relatively well identified, a wider range of exposure assignment p may lead to similar estimations. The distribution of the segregation index (as the average similarity of peers) shows that very different spatial patterns can lead to the same interpretation of neighborhood effects: not only are the estimations heavily biased, they do not describe the system sufficiently well. What is interesting to note

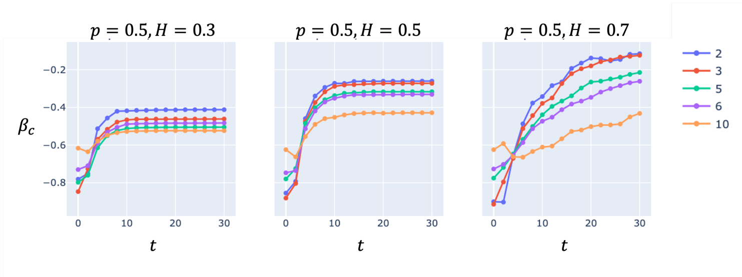
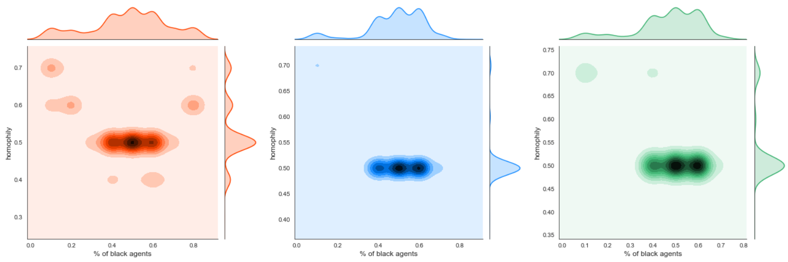
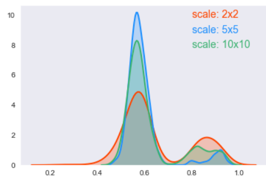


Figure 2 Variation in estimation of OLS estimator for color effects (β_c) under misspecification of neighborhood for resp. (2x2), (3x3), (5x5), (6x6), (10x10) area grids.



(a) 2x2 **(b)** 5x5 **(c)** 10x10



(d) segregation index

Figure 3 Posterior distributions for H and p approximated from the ABC and the estimation of color effect for different scales (a-c) ; Distribution of segregation in the selected simulations for scales 2x2, 5x5 and 10x10 (d).

is the influence of scale on the uncertainty. It appears there exists a scale for which the possibilities are reduced and the posterior distributions are more concentrated (for instance see Fig. 3.b))

5 Discussion

This very simple model was used to illustrate the use of ABM pattern oriented methods to question the uncertainty of causal model estimates. The specific representation of space, here as neighborhood effects, will have an impact on the meaning of the estimands and ultimately, on the appropriate interpretation of estimations. Notions of spatial equifinality in complex systems should be considered to better understand the role of space in social mechanisms. A relevant road-map for spatial causal inference would consider the specific challenges that relate to defining this spatial context. Thinking of causality within a spatial context may imply moving from a paradigm of causal dependence between variables to one of causal mechanisms in a system of spatially embedded agents.

References

1 Steven N Durlauf. Neighborhood effects. *Handbook of regional and urban economics*, 4:2173–2242, 2004.

2 Scott Ferson and Kari Sentz. Epistemic uncertainty in agent-based modeling. In *7th international workshop on reliable engineering computing*, pages 65–82, 2016.

3 A Stewart Fotheringham and Mehak Sachdeva. Modelling spatial processes in quantitative human geography. *Annals of GIS*, 28(1):5–14, 2022.

4 George C Galster. The mechanism (s) of neighbourhood effects: Theory, evidence, and policy implications. In *Neighbourhood effects research: New perspectives*, pages 23–56. Springer, 2011.

5 Jackie Kazil, David Masad, and Andrew Crooks. Utilizing python for agent-based modeling: The mesa framework. In Robert Thomson, Halil Bisgin, Christopher Dancy, Ayaz Hyder, and Muhammad Hussain, editors, *Social, Cultural, and Behavioral Modeling*, pages 308–317, Cham, 2020. Springer International Publishing.

6 David Manley, Robin Flowerdew, and David Steel. Scales, levels and processes: Studying spatial patterns of british census variables. *Computers, environment and urban systems*, 30(2):143–160, 2006.

7 Josie McCulloch, Jiaqi Ge, Jonathan A Ward, Alison Heppenstall, J Gareth Polhill, and Nick Malleson. Calibrating agent-based models using uncertainty quantification methods. *Journal of Artificial Societies and Social Simulation*, 25(2), 2022.

8 J Michael Oakes. The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social science & medicine*, 58(10):1929–1952, 2004.

9 Francis A Pearman. Gentrification and academic achievement: A review of recent research. *Review of Educational Research*, 89(1):125–165, 2019.

10 Brian J Reich, Shu Yang, Yawen Guan, Andrew B Giffin, Matthew J Miller, and Ana Rappold. A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review*, 89(3):605–634, 2021.

11 Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

12 Flaminio Squazzoni. The impact of agent-based models in the social sciences after 15 years of incursion. *The Impact of Agent-Based Models in the Social Sciences after 15 Years of Incursion*, pages 1000–1037, 2010.

13 Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

14 Brandon M Turner and Trisha Van Zandt. A tutorial on approximate bayesian computation. *Journal of Mathematical Psychology*, 56(2):69–85, 2012.

15 Konstantina Valogianni and Balaji Padmanabhan. Causal abms: Learning plausible causal models using agent-based modeling. In *The KDD'22 Workshop on Causal Discovery*, pages 3–29. PMLR, 2022.

16 Maarten Van Ham, David Manley, Nick Bailey, Ludi Simpson, and Duncan Maclennan. Neighbourhood effects research: New perspectives. In *Neighbourhood effects research: New perspectives*, pages 1–21. Springer, 2011.

17 Blair Wheaton, Rosane Nisenbaum, Richard H Glazier, James R Dunn, Catharine Chambers, et al. The neighbourhood effects on health and well-being (nehw) study. *Health & place*, 31:65–74, 2015.