Quantitative methods I: Reckoning with uncertainty

Rachel Franklin
Centre for Urban and Regional Development Studies (CURDS), Newcastle University, Newcastle Upon Tyne, UK;
The Alan Turing Institute, London, UK

Abstract
In this report, I address the timeless topic of uncertainty. Quantitative methods are in flux and uncertainty provides a useful vantage point for assessing these changes. I draw particular attention to two forces at odds with each other: on the one hand, innovative methods and data that allow researchers to embrace uncertainty, and on the other, recent trends towards deliberate amplification of uncertainty, in the form of differential privacy methods which potentially limit our capacity for measuring and understanding the world around us. I close by highlighting some implications for users and producers of quantitative methods and data in geography.

Keywords
uncertainty, geography, quantitative methods, big data, differential privacy

I Introduction
If there is one certainty in quantitative methods, it is that uncertainty is always with us. Hidden behind the analytical curtains, hard-wired into data collection and interpretation, and fundamental to both methodological and conceptual development, uncertainty shapes every stage of the research process in quantitative human geography. Indeed, the existence of uncertainty – here defined broadly as the gap that exists between real-world, ‘true but unknown’ values and relationships and what we are able to observe, given the methods and data available – is our raison d’être as researchers; were we able to observe true but unknown values directly, there would, after all, be no need for further investigation.

Of course, quantitative methods are not unique in this respect; from assorted perspectives and positions, all of geography wrestles in some fashion with uncertainty (Fusco et al., 2017). This universality is, in fact, one impetus for the focus on uncertainty in this report on Quantitative Methods. The ways in which quantitative methods approach uncertainty, and how it is implicated in contemporary advances and challenges, are meaningful not only to this particular corner of the discipline, but to all of geography.

If uncertainty is a constant in quantitative methods, why write about it now? The answer is that quantitative methods are in flux and one way of comprehending the changes that are occurring is through an uncertainty lens. For one thing, the range of methods employed in quantitative geography is
undergoing rapid expansion, due to growth in data types and provenance but also related to subtle changes in the types of research done in the field. This expansion both maintains traditional relationships with uncertainty but also introduces new, important ways in which it matters.

Second, where arguably many traditional quantitative approaches subsumed uncertainty under a veneer of normative truth – what Poon (2005) refers to as ‘the monopoly of logical positivism as the central way of knowing’ – newer methods explicitly address and incorporate the uncertainty of reality. This is exciting research that bridges the conceptual and the methodological, often leveraging increased availability of high-resolution spatial data. Continued innovation is highly contingent on the availability of certain kinds of data.

However, data are also changing. This is, of course, well known where ‘big data’ are concerned (McAfee et al., 2012). This report focuses on other landmark shifts that are also occurring, as traditional data providers (such as governments) make existing uncertainty estimates more visible and, more recently, as they wrestle with new forms of strategic injections of uncertainty into data as a way of maintaining respondent confidentiality, whilst still aiming to provide high-quality data as inputs not only to research but also policymaking, election redistricting and local-area funding mechanisms. These efforts have particularly strong ramifications for geographers and others who depend on small-area data (such as census tracts or output areas). Uncertainty has never been more important.

In this first of three reports on quantitative research methods, all organized around the themes of flux and continuity in quantitative methods,¹ I outline the bedrock role of uncertainty in quantitative methods, emphasizing some principal elements. I then turn to recent research that specifically aims to reckon with uncertainty and reflect on the implications for the data we use, and potential challenges on the horizon.

II Uncertain foundations

To understand where we are, it helps to look at how we got here. Much has already been written about uncertainty in quantitative methods – it is an evergreen subject. This reflects its importance, but the range of perspectives adopted also underscores that it means different things to different people (Fusco et al., 2017). One important distinction is that made by Derbyshire (2020), who highlights the difference between epistemological uncertainty, or ‘the accuracy of what we know presently’, and ontological uncertainty, which, in my own paraphrasing, can be thought of as changes to the ‘true but unknown’ world. Most commentary on uncertainty in quantitative methods, including this report, is focused on the former – the factors that create a gap between what we observe or model, and the actual phenomena we are studying. In GIScience, this includes concerns with error propagation (Heuvelink, 2002) and spatial information (Goodchild, 2018). Griffith (2018), one of many to list and classify potential sources of uncertainty, identifies the following: calculation, measurement, specification, sampling and stochastic. Wei and Murray (2012), in discussing uncertainty in spatial optimization methods, identify many of the same sources, but focus on the need to explicitly account for and measure uncertainty in existing models. From yet another standpoint, in this case, physical geography and environmental science, Brown (2010, p. 77) states that uncertainty can be defined as ‘a state of confidence. Here, confidence is defined in the broadest sense of (degree of) trust or conviction in knowledge, which includes the narrower sense of “statistical confidence”’. This is a helpful definition, which stresses that uncertainty is about much more than statistics or models, and, in its essence, comes down to confidence in our results, as quantitative human geography researchers. The range of definitions of uncertainty is also important. It signifies that, although we may all arrive at the same destination – the central role of uncertainty in quantitative methods – many have travelled different routes to reach it.

Coming to terms with uncertainty, if we concur that it has many guises and impacts research in a variety of ways, can be difficult. O’Sullivan (2004), writing about complexity and human geography, suggests that acknowledging the (deterministic) uncertainty that arises from the unpredictability of systems highlights ‘the futility of prediction’ (p. 283) but also opens up new possibilities for models as ‘thought experiments’ or narrative tools. Stepping...
back and surveying the whole of quantitative methods, this insight suggests that uncertainty in our methods is not a deal-breaker, but rather a tool that can help produce unexpected insights. Or, in common parlance, uncertainty is a feature, not a bug.

Adopting Brown’s (2010) definition of uncertainty – a state of confidence in knowledge – some interesting components of uncertainty in quantitative methods emerge. Although they may not be as visible to those outside the field, these are topics that, among quantitative human geographers, animate discussion about reliability of research findings. Among them are: statistical uncertainty, sampling and representation uncertainty, construct uncertainty and the modifiable areal unit problem (which, like uncertainty, seems to always be with us).

The most obvious of these is statistical uncertainty. By their very nature, inferential statistics are about measuring confidence in knowledge; models attempt to approximate reality, but as abstractions they are never perfect. (And mainstream machine-learning methods are little better; classic estimates of error or uncertainty may be absent, but that does not mean the uncertainty is, too.) A still-common criticism of geographical models and modellers is that they tend to treat analytical results as truth. In reality, quantitative geography has long moved away from pre-ordained. This, in turn, implies that characterization systems in the world today, it is no less constructed than the Facebook advertising platform or, say, the Golden Gate Bridge… all these structures were created by people: people living in a particular place, at a particular time, and who were influenced – as we all are – by the world around them’. Where race and ethnicity categories are concerned, this shortcoming is well known (Mateos et al., 2009); censuses, administrative data and surveys may allow for self-identification of race, for example, but the categories from which respondents must choose are not pre-ordained. This, in turn, implies that characteristics such as population composition, diversity, or segregation have a degree of uncertainty embedded in them, independent of choice of analytical method. The same holds for constructs like migration and, increasingly, for those we may have previously taken for granted, such as employment or occupation.

Often researchers have to grapple with applications that concern not only individuals, but also areal aggregates. This deepens the potential for uncertainty: not only from validity of inputs, but sample size, methodological approach, and the modifiable areal unit problem (MAUP), as well. MAUP is...
defined as a boundary and a zonation challenge (Openshaw and Taylor, 1981) but can be summarized in lay terms as, ‘the choice of spatial units may affect results’. Ideally, researchers match spatial unit to the process under investigation; the distance between what is measured and what is actually occurring is minimized. In reality, researchers typically adopt the spatial units that are available and best match the hypothesized scale of the phenomenon being studied. The outcome is that a degree of uncertainty adheres to the findings.

III Uncertain methods

Recent research in quantitative human geography has capitalized on uncertainty and the ways in which it confounds our capacity to model the human experience – uncertainty as a feature. Kwan’s (2012, 2018) contribution to our understanding of the ‘uncertain geographic context problem’ is an example of such research, but so is Brunsdon, Fotheringham, and Charlton’s (1998) Geographically Weighted Regression (GWR), which acknowledges the uncertainty that underlies global regression estimates that presume spatial stationarity between determinants and outcomes. Indeed, over the past decade or so, uncertainty challenges related to spatial scale and context have been at the forefront of quantitative methods development. Representative examples in my own population subfield from a very large literature include: Reardon et al. (2008) and Fowler (2015) on multi-scalar segregation profiles, Clark et al. (2015) on segregation and diversity, the measurement of neighbourhood context (Andersson and Malmberg, 2014), and neighbourhood definition (Spielman and Logan, 2013).

A flurry of recent research expands on the quantification of uncertain exposure, context and neighbourhood, building off the consensus that administrative units such as census tracts, blocks, or output areas are insufficient proxies or containers for actual, spatial lived experience. Rather than amalgamations of units, buffers, or comparative analysis across multiple spatial scales, newer research pairs very high-resolution data with multi-scalar methods to more clearly home in on hypothesized neighbourhood or contextual effects. These methods do not necessarily depend on ‘big data’ in the sense that we have come to understand it, but rather ‘sensitive data’ in that location, movements and temporal variations are finely measured.

In fact, the lead role of the data is what distinguishes much of the recent innovative research working with uncertain contexts – or, rather, data and methods have been cast as dual, highly co-dependent, leads. Victoriano, Paez, and Carrasco (2020), for example, use machine-learning methods to characterize mobility strategies. Their seemingly small set of participants (165) belies the complexity of the data: a 7-day travel diary for each results in 1128 days of data and over 16,000 trips and activities. Fowler et al. (2020) address contextual effects in segregation – in many ways similar to research contributions noted above, except that they rely on secure access to individual-level census data for their analysis. Their emphasis is on what they term the ‘contextual fallacy’, the extent to which individuals in the same spatial unit (e.g. census tract) have different contextual experiences. In a similar vein, Petrović et al., (2018, 2021) estimate multiscale contextual effects, but starting from individuals located at a resolution of 100 m by 100 m. Pearce (2018), on health and exposure, expands the conceptual uncertainty bounds to include both space and time, arguing that exposure accumulates over the life course, which is a function of both time (years lived), but also space (residential stability and mobility). As Pearce highlights, to more accurately account for potential exposure over the life course requires data on individual trajectories over very long timespans, but also environmental data over time and at sufficient spatial resolutions to be able to account for individual context. This is not primarily a methodological challenge, but a data challenge. Folch, Fowler and Mikaelian (2021) also consider how uncertainty in context and exposure can be measured, estimating air and water toxicity and child mobility, both over time and from day to day (home versus daycare location). Their analysis explicitly addresses questions around uncertainty: positional accuracy of children, relative to toxicity measures, neighbourhood context and child mobility.

This is only one narrow slice of the data-method innovation nexus, but of course similar revolutions are occurring across the quantitative methods spectrum. Increased availability of location data, whether
from devices or satellites, vastly expands our capacity for modelling the real world, which in turn demands novel methods for bridging theory and data. Uncertainty is a leitmotif: highly certain locational attributes, likely increased uncertainty on other dimensions such as representativeness, and the possibility of new methods that help render visible the uncertainties that surround so much of human behaviour, interaction and systems.

The tension between methodological advances, uncertainty and data requirements is nicely encapsulated in Petrović et al., (2020) in this journal. Speculating about what sorts of data would be necessary to test existing theories about the importance of neighbourhood effects for a range of outcomes, they land on the importance of quantitative methods – multi-scalar measures – but also microgeographic data. And herein lies the potential problem: contemporary quantitative methods permit more nuanced and sophisticated understanding of a range of geographic and social phenomena, thereby hopefully decreasing uncertainty, or increasing our confidence in knowledge. In parallel, however, the availability of high-resolution data required for these methods increasingly runs against the grain of heightened perception of possible loss of privacy and confidentiality. One solution to these concerns is to shift uncertainty onto the data via differential privacy tools, thereby protecting individuals. This deliberate insertion of error into existing data products has repercussions not only for quantitative methods, but all of geography and a range of civic stakeholders and policymakers, as well.

IV Uncertain data

There is a recent precedent for uncertain data to disrupt geographical research. In the United States, the bread and butter of geographical analysis has long been U.S. Census data products, whether microdata, reflecting individual characteristics, or summary files for a range of geographical units, from states to counties to census tracts, block groups and blocks. Until 2010, research relied on short form (full census counts but for limited variables) and long form (sample data on a range of questions, including migration, educational attainment and income). Uncertainty was a factor for these estimates, even for the full count data – undercounts have always presented a challenge for particular sub-groups and places. Data tables for geographic areas presented only point estimates, however, and no margins of error (MOEs), even for data based on long form sample data. Thus, although researchers were in theory aware of data limitations, in practice the data were often treated as fact. This changed with the advent of the American Community Survey (ACS), which came fully online in 2005 and entirely replaced the long form for the 2010 decennial census. The ACS offers advantages over the long form: rather than providing important socio-economic updates once a decade, the ACS is a rolling survey that goes out to a sample of households every month. This provides timelier information and, for researchers, has helped ensure that analysis is not wildly out of date by the time results are published. However, the ACS sample size is considerably smaller than that for the long form and this has had two important impacts.

First, for the first time, MOEs were published alongside estimates, disclosing to many researchers the very unstable (read: uncertain) ground upon which they were conducting their neighbourhood and local analyses. The visibility of uncertainty was problematic – should researchers and local stakeholders simply ignore the MOEs? As Jurjevich et al. (2018) have shown, many users of ACS data, including local planners and stakeholders, simply do not know how to interpret the uncertainty embedded in MOEs. Practically speaking, wider margins of error mean more uncertainty, so that it can be difficult to know what the ‘true’ neighbourhood characteristics of a place actually are. Spielman and Singleton (2015) propose composite or geodemographic measures for neighbourhoods that combine characteristics to provide more certainty for local areas, but this is a limited solution.

Second, and most importantly for quantitative spatial researchers, ACS uncertainty is not constant across space – some areas have higher uncertainty in estimates than others. As Folch et al. (2016) document, uncertainty varies both locally and regionally and, crucially, appears to be related to the characteristics of places (more uncertainty in lower-income
areas, for example). Whilst this has clear implications for those working with the full universe of areas, it can also affect basic understanding of individual places: ‘For example, in census tract 190,602 in the Belmont Cragin neighbourhood of Chicago, Illinois, the number of children in poverty is somewhere between 9 and 965 (2006–2010 ACS estimates)’ (Folch et al., 2016: p. 1537).

My point here is that methodological advances do not exist in a vacuum, but rather in concert with changes to data infrastructure. The example given here is from the United States but is not likely to be an isolated case, given evolving conversations around data and privacy in the age of big data and fast-growing private and public surveillance apparatuses. Whether data become newly uncertain or simply have a light shone on their pre-existing uncertainty (in the ACS it was sadly both), this impacts both the types of analytical advances that can be expected and the confidence we as researchers can have in our results. Moreover, where data and uncertainty are concerned, this is in many ways a best-case scenario. Government providers have strict protocols for quality assessment; other data providers have no such responsibility to make uncertainty visible or prominent in their offerings.

V Uncertainty is dead; long live uncertainty

It is one thing to confront the imperfections of existing data sources, such as the American Community Survey. It is another entirely to face the prospect of data deliberately rendered more uncertain, and yet that is where we are. Quantitative methodological innovation works in tandem with increased sophistication of data, both ‘big’ and ‘sensitive’. With these innovations comes increased risk of violation of individual privacy and confidentiality, as they are understood (and socially constructed) in today’s society. This is a particular challenge for government data providers, many of whom have a legal obligation to guarantee confidentiality. Either our expectation of privacy must evolve or the data will have to.

The test case for this new reality is, again, the US Census Bureau. With the 2020 Census, the Bureau will introduce additional error to all statistics for areas below the state level, in a process termed differential privacy (Ruggles and Van Riper, 2021). This represents a major shift in data provision. As Hawes (2020) puts it, ‘Consumers of official statistics, particularly those who use data products that have been produced for a long time, are accustomed to the data looking a certain way, and to interpreting those data as the “ground truth.” As such, they are unaccustomed to seeing population counts with fractional or negative values’.

Quantitative geographers may be among the most impacted researchers, given their focus on aggregate geographic data, but the Census Bureau Disclosure Avoidance System and differential privacy are likely to have wide-reaching effects for non-quantitative researchers and policymakers, as well. Preliminary research has suggested that uncertainty may be much higher for certain places and groups, for example, indicating decadal population loss when none has occurred, for small towns and indigenous areas (Wezerek and Van Riper, 2020) or mis-characterizing population counts and characteristics in political redistricting (Kenny et al., 2021). In estimating possible effects on county-to-county migration data, Winkler et al. (2021) find that uncertainty is potentially higher for Hispanic migrants, as well as the young and old. Smaller-population counties, rural areas and the Great Plains of the United States may also be differentially affected by more uncertain data. The effects of differential privacy on data provision may be widespread and pernicious. Evaluating the impacts of the COVID-19 pandemic on mortality rates, Hauer and Santos-Lozada (2021) emphasize the importance of reliable denominators – the source of which is often Census Bureau data – in constructing age-specific mortality rates.

The Census Bureau differential privacy example is just one front on a wide-ranging debate about the high-resolution, fine-grained data that nourish the development of methods and theory in quantitative human geography, but it may be a harbinger of things to come. Although the preliminary assessments of the Bureau’s algorithm suggest that the risk of disclosure is no higher than what would be expected at random (Ruggles and Van Riper, 2021), the direction of the prevailing wind is clear: the days of (relatively)
easy access to reliable high-resolution data may be limited. And whilst this may be a boon to individual privacy, the downsides are also evident: not only increased uncertainty in data, which hampers development of quantitative methods and knowledge, but also – quite likely – an entrenchment of already-unequal privileged access to high-quality data, and a reinforcement of societal inequities that mean some groups and places are better measured (and understood) than others.

VI Conclusions: With great data comes great responsibility

Research in quantitative human geography has blossomed over the past several years, as better data and computational ease have facilitated engagement with tricky questions that, theoretically anyway, have long entailed a high degree of uncertainty. This report has scarcely scratched the surface of dynamic and conceptually rich research currently published across the continuum of quantitative human geography. And yet, as this report has shown, new data-related dilemmas are emerging which, although they may not completely disrupt innovation, may very well introduce new forms of uncertainty. For example, big data require curation in order to be easily usable and the data engineering that underpins this curation is often opaque where it should be transparent (Arribas-Bel et al., 2021). In addition, where quantitative researchers have traditionally been consumers of data products, recent method developments, especially on uncertain context, indicate that we may soon be producers of bespoke geographies, such as neighbourhoods. How will we make clear the uncertainties embedded in these geographies?

Equally importantly, how can we contribute to conversations that help disentangle the needs of the few and the needs of the many, where data are concerned? High-quality data provision is not only about arcane model development and researcher privilege; it is also about research that feeds equitable policy development and visibility of under-represented groups. As D’ignazio and Klein (2020) emphasize: ‘What gets counted counts’. Uncertainty is an intrinsic component of quantitative research, but new emphasis on differential privacy methods – although laudable from an individual privacy perspective – poses very real risks not only for geographical research but also for disadvantaged groups and places who rely on accurate numbers and statistics as a form of representation.

Acknowledgments
One certainty is the importance of thoughtful and constructive colleagues. This report benefited enormously from comments on earlier drafts from Daniel Arribas-Bel, Steven Manson, Antonio Páez, and Caitlin Robinson.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Rachel Franklin https://orcid.org/0000-0002-2614-4665

Notes
1. Although, in keeping with the theme of this report, there is uncertainty in my plans. Consider this footnote a ‘pre-registration’ of report intentions.
2. For example, the long form of the Census asked a 5-year migration question, measuring moves that were already up to 5 years old at the time of enumeration. Those working with migration data at the cusp of the following Census (e.g. researching in 2010, but with 2000 Census data), were partly measuring moves that took place 15 years beforehand (Franklin and Plane, 2006).
3. As Kenny et al. (2021) point out, the US Census Bureau has long adjusted data (e.g. record swapping) to help ensure confidentiality, however, the new Disclosure Avoidance System, which combines injecting noise into data with post-processing adjustments, is widely agreed to be an entirely different sort of confidentiality protection beast.

References
Andersson EK and Malmberg B (2014) Contextual effects on educational attainment in individualised, scalable
neighbourhoods: differences across gender and social class. Urban Studies 52(12): 2117–2133.
Arribas-Bel D, Green M, Rowe F, et al. (2021) Open data products - a framework for creating valuable analysis ready data, journal of geographical systems (in press)
Brown JD (2010) Prospects for the open treatment of uncertainty in environmental research. Progress in Physical Geography: Earth and Environment 34(1): 75–100. DOI: 10.1177/0309133309357000.
Brunsdon C, Fotheringham S and Charlton M (1998) Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician) 47(3): 431–443.
Clark WA V., Anderson E, Osth J, et al. (2015) A multiscalar analysis of neighborhood composition in Los Angeles, 2000-2010: a location-based approach to segregation and diversity. Annals of the Association of American Geographers 105(6): 1260–1284. DOI: 10.1080/00045608.2015.1072790.
Derbyshire J (2020) Answers to questions on uncertainty in geography: old lessons and new scenario tools. Environment and Planning A: Economy and Space 52(4): 710–727. DOI: 10.1177/0308518X19877885.
D’ignazio C and Klein LF (2020) “What gets counted counts.” in data feminism. Retrieved from: https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp.
Folch DC, Fowler CS and Mikaelian L (2021) Day time, night time, over time: geographic and temporal uncertainty when linking event and contextual data. Environmental Health: A Global Access Science Source 20(1): 51.
Folch DC, Arribas-Bel D, Koschinsky J, et al. (2016) Spatial variation in the quality of American Community Survey estimates. Demography 53(5): 1535–1554.
Fowler C (2015) Segregation as a multiscalar phenomenon and its implications for neighborhood-scale research: the case of South Seattle 1990-2010. Urban Geography 37(1): 1–25.
Fowler CS, Frey N, Folch DC, et al. (2020) Who are the people in my neighborhood?: The “contextual fallacy” of measuring individual context with census geographies. Geographical Analysis 52(2): 155–168.
Franklin RS and Plane DA (2006) Pandora’s box: the potential and peril of migration data from the American community survey. International Regional Science Review 29(3): 231–246. DOI: 10.1177/0160017606289895.
Fusco G, Caglioni M, Emsellem K, et al. (2017) Questions of uncertainty in geography. Environment and Planning A: Economy and Space 49(10): 2261–2280. DOI: 10.1177/0308518X17718838.
Goodchild M. F. (2018) A GIScience perspective on the uncertainty of context. Annals of the Association of American Geographers 108(6): 1476–1481.
Graham M and Shelton T (2013) Geography and the future of big data, big data and the future of geography. Dialogues in Human Geography 3(3): 255–261. DOI: 10.1177/2043820613513121.
Griffith DA (2018) Uncertainty and context in geography and giscience: reflections on spatial auto-correlation, spatial sampling, and health data. Annals of the American Association of Geographers 108(6): 1499–1505.
Hauer ME and Santos-Lozada AR (2021) Differential privacy in the 2020 Census will distort COVID-19 rates. Socius 7: 2378023121994014.
Hawes MB (2020) Implementing differential privacy: seven lessons from the 2020 United States Census. Harvard Data Science Review 2(2).
Heuvelink GB (2002) Analysing uncertainty propagation in GIS: why is it not that simple. Uncertainty in Remote Sensing and GIS: 155–165.
Jurjevich JR, Griffin AL, Spielman SE, et al. (2018) Navigating statistical uncertainty: how urban and regional planners understand and work with American community survey (ACS) data for guiding policy. Journal of the American Planning Association 84(2): 112–126.
Kenny CT, Kuriwaki S, McCartan C, et al. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 US Census. Science advances 7(41).
Kwan MP (2012) The uncertain geographic context problem. Annals of the Association of American Geographers 102(5): 958–968.
Kwan MP (2018) The limits of the neighborhood effect: Contextual uncertainties in geographic, environmental health, and social science research. Annals of the American Association of Geographers 108(6): 1482–1490.
Mateos P, Singleton A and Longley P (2009) Uncertainty in the analysis of ethnicity classifications: issues of extent and aggregation of ethnic groups. Journal of Ethnic and Migration Studies 35(9): 1437–1460. DOI: 10.1080/13691830903125919.
McAfee A, Brynjolfsson E, Davenport TH, et al. (2012) Big data: the management revolution. Harvard Business Review 90(10): 60–128.

Openshaw S and Taylor P (1981) Quantitative Geography A British View. In: Wrigley N and Bennett R (eds), The Modifiable Areal Unit Problem. London: Routledge.

O’Sullivan D (2004) Complexity science and human geography. Transactions of the Institute of British Geographers 29(3): 282–295. http://www.jstor.org/stable/3804492.

Petrović A, Manley D and van Ham M (2020) Freedom from the tyranny of neighbourhood: Rethinking sociospatial context effects. Public Health Genomics 44(6): 1103–1123.

Petrović A, van Ham M and Manley D (2018) Multiscale measures of population: Within- and between-city variation in exposure to the sociospatial context. Annals of the American Association of Geographers 108(6): 1491–1498. DOI: 10.1080/24694452.2017.1416280.

Petrović A, van Ham M and Manley D (2021) Where do neighborhood effects end? moving to multiscale spatial contextual effects. Annals of the American Association of Geographers: 1–21.

Phillips JD (2004) Doing justice to the law. Annals of the Association of American Geographers 94(2): 290–293. DOI: 10.1111/j.1467-8306.2004.09402006.x.

Poon JPH (2004) Quantitative methods: past and present. Public Health Genomics 28(6): 807–814. DOI: 10.1191/0309132504ph521pr.

Poon JPH (2005) Quantitative methods: not positively positivist. Public Health Genomics 29(6): 766–772. DOI: 10.1191/0309132505ph583pr.

Reardon SF, Matthews SA, O’Sullivan D, et al. (2008) The geographic scale of metropolitan racial segregation. Demography 45(3): 489–514. DOI: 10.1353/dem.0.0019.

Robbin A (1999) The problematic status of u.s. statistics on race and ethnicity. Journal of Government Information 26(5): 467–483.

Spielman SE and Logan JR (2013) Using high-resolution population data to identify neighborhoods and establish their boundaries. Annals of the Association of American Geographers 103(1): 67–84. DOI: 10.1080/00045608.2012.685049.

Spielman SE and Singleton A (2015) Studying neighborhoods using uncertain data from the American community survey: a contextual approach. Annals of the Association of American Geographers 105(5): 1003–1025.

Victoriano R, Paez A and Carrasco JA (2020) Time, space, money, and social interaction: Using machine learning to classify people’s mobility strategies through four key dimensions. Travel Behaviour and Society 20: 1–11.pp.

Wei R and Murray AT (2012) An integrated approach for addressing geographic uncertainty in spatial optimization. International Journal of Geographical Information Science 26(7): 1231–1249. DOI: 10.1080/13658816.2011.633918.

Wezerek G and Van Riper D 2020. Changes to the census could make small towns disappear. The New York Times.

Winkler RL, Butler JL, Curtis KJ, et al. (2021) Differential privacy and the accuracy of county-level net migration estimates. Population Research and Policy Review: 1–19.