



Wolf, L. J., Fox, S., Harris, R., Johnston, R., Jones, K., Manley, D., Tranos, E., & Wang, W. W. (2021). Quantitative geography III: Future challenges and challenging futures. *Progress in Human Geography*, 45(3), 596-608. <https://doi.org/10.1177/0309132520924722>

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC

Link to published version (if available):
[10.1177/0309132520924722](https://doi.org/10.1177/0309132520924722)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Sage at <https://doi.org/10.1177/0309132520924722>. Please refer to any applicable terms of use of the publisher.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>



Quantitative geography III: Future challenges and challenging futures

Progress in Human Geography
2021, Vol. 45(3) 596–608
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0309132520924722
journals.sagepub.com/home/phg



Levi John Wolf , Sean Fox, Rich Harris, Ron Johnston , Kelvyn Jones, David Manley, Emmanouil Tranos , and Wenfei Winnie Wang

University of Bristol, UK

Abstract

In the previous two reports in this series, we discussed the history and current status of quantitative geography. In this final report, we focus on the future. We argue that quantitative geographers are most helpful when we can simplify difficult problems using our distinct domain expertise. To do this, we must clarify the theory underpinning core conceptual problems in quantitative geography. Then, we examine the social forces that are shaping the future of quantitative geography. We conclude with criteria for how quantitative geography might succeed in addressing these challenges.

Keywords

causality, data science, inclusiveness, modifiable areal unit problem, open science, quantitative geography, replication, spatial dependence

I Introduction

The future of quantitative geography lies in the comparative advantage of its theory. If it's the case that 'by our theories you shall know us' (Harvey, 1969: 486), we should ask 'what unique output or consequences occur due to spatial and geographic thinking?' (Golledge, 2002: 11). Searching questions like this were common in the early 2000s as quantitative geographers sought to mobilize the field for a new millennium. Yet, they remain difficult to answer about the past 20 years of quantitative geography, let alone the next.

Unfortunately, predictions about the past are less useful than predictions about the future. To clarify the future of quantitative geography, it will help to first discuss the core theory of quantitative geography. Then, we discuss current

forces that act upon this core. We will conclude with a sketch of what success might look like.

II Theoretical forces driving quantitative geography

It is true, as in the case of other ossifications, that attacking this ossification is almost sure to reduce the apparent neatness of our subject. But neatness does not accompany rapid growth. (Tukey, 1962: 8)

Corresponding author:

Levi John Wolf, University of Bristol School of Geographical Sciences, University Road, Bristol BS8 1SS, UK.

Email: levi.john.wolf@bristol.ac.uk

To understand what quantitative geography may offer in the future, we need to take stock. Accumulation of knowledge in quantitative geography is a cyclical process: waves of scholars converge on similar empirical and computational challenges, then follow their own rivulets back to common conceptual understanding. Historical shifts in perspective are tidal; geography is transformed both by individual geographers' mercurial interests as well as shifts in socially or scientifically-relevant questions (Johnston and Sidaway, 2015). Beyond continual change, it is erosion of complexity – slowly and progressively by accumulated knowledge – that is the point.

Thus, we present geographical theories as *heuristics* that quantitative geographers use to simplify reality. Fortunately, 'accumulated' knowledge in quantitative geography involves only a few heuristics (Goodchild, 2004). Previous reports in this series have traced them through time. Here, we define their main points.

I Obeying the law is not the point of the law

Spatial dependence is often explained in terms of Tobler's Law:

Everything is related to everything else, but near things are more related than distant things. (Tobler, 1970: 236)

Claims about the 'death of distance' and the digital economy (Cairncross, 2001) fostered more critical attention to the definition of 'near' (Miller, 2004), but Tobler's Law is flexible enough to still be meaningful after this redefinition (e.g. Hecht and Moxley, 2009; Tranos and Nijkamp, 2013). When Miller (2004) bolsters Tobler's Law with 'near is a more flexible and powerful concept than commonly appreciated' (p. 285), he is right.

But, he also bends the meaning of 'law' (Smith, 2004). In this definition, Tobler's Law is too vague for most of social science (Merton, 1949). Further, as Gibbons and Overman (2012)

argue, our knowledge about geographical processes probably will not benefit from exploring the many notions of 'near' that may support Tobler's Law. The impact of dependence implied by Tobler's Law often stays the same (LeSage and Pace, 2014) and is rarely the point of an empirical model.

Instead, Tobler's Law should be used as Waldo Tobler (1970) intended: to simplify problems. In this spirit, Church (2018) uses the Law to simplify warehouse location problems. Classical methods consider *all* warehouses when supplying *each* demand site, so odd decisions are considered but never made, like Detroit, Michigan, supplying Bakersfield, California. Church (2018) only explicitly models 'near' facilities, and finds optimal solutions by gradually increasing what is 'near' each facility. This is Tobler's Law enforced, not obeyed. Future work should pursue simplification in this manner, too.

2 MAUP and the challenge of uncertainty laws

The *Modifiable Areal Unit Problem* (Openshaw, 1984), and its recent relatives (Kwan, 2018), reflect another kernel of geographical knowledge. The MAUP implies separate 'scale' and 'zoning' problems (Fotheringham et al., 2000). Neither helps us simplify analyses, but each clarifies assumptions we must make. The 'zoning' problem implies:

Place characteristics are contingent on where place boundaries are drawn.

Alternatively, the 'scale' problem echoes Broido and Clauset's (2019) result about the rarity of 'scale-free' networks:

Characteristics of a geographical process are contingent on the scale at which the process is measured.

Interactions between boundaries, scale, place, and process are complex, so assumptions

are required to avoid the MAUP. Fortunately, Fotheringham and Wong's (1991) three methods to resolve the MAUP outright remain helpful. The first, 'frame-independent' analysis, has vexed geographers (cf. Tobler, 1989). Indeed, King (1996) argues scale sensitivity can be theoretically important. Thus, we focus on the remaining two: drawing better areas and moving beyond the area.

2.i Building a better MAUP trap

Optimal zoning methods have long shown promise for solving the MAUP. Optimal zoning methods estimate boundaries/places given assumptions about how they should look. This usually solves the 'zoning' problem, and the 'scale' component is solved separately. Promising work provides mathematical proof of minimum-error aggregations (Bradley et al., 2016), rigorously characterizes the sensitivity of statistics to the MAUP (Duque et al., 2018), builds zones with less internal uncertainty (Spielman and Folch, 2015), or develops general-purpose zones from demographic data (Johnston et al., 2004; Singleton and Spielman, 2014). Zoning algorithms have improved remarkably (Tam Cho, 2018), so this remains promising.

Beyond an 'optimal' zoning, we might analyse random samples from a distribution of zoning systems (Tapp, 2019). This would liberate us from any single zoning and provide tests on how unusual observed zones are. However, we must be cautious: Tam Cho and Liu (2018) show that failure to sample *uniformly* from the unfathomably many possible zoning systems will unpredictably bias even simple statistics.

2.ii Empirical answers to theoretical questions

One source of optimal zoning, Isard's (1956) ruminations about 'true' regions (p. 17), illustrates a tension from Jones (1998). One perspective on scale and zoning is that they define an

epistemological frame – a hierarchy in which geographical processes 'make sense'. Another is ontological, with scale and zoning describing a system of relationships from which independently meaningful contexts emerge. This ontological perspective grounds the half-century search for 'conchorations' (Nelson, 2020), where regions found in different processes converge to common boundaries (Isard, 1956: 20).

But, convergence itself does not map useful theoretical concepts like 'context' or 'exposure' onto empirical zones (King, 1996; Kwan, 2018). This means optimal zoning offers 'empirical answers to theoretical questions' (King, 1996: 160). Optimal zones themselves are atheoretical, with no direct link to a specific process or theory about place, context, or exposure. As such, they only 'fix' our frame of analysis. They cannot tell us if estimates about this frame make sense.

Therefore, we must accept that the MAUP 'is not an empirical problem; it is a theoretical problem' (King, 1996: 163). Re-aggregation is a theoretical act with empirical effects. A new theory about the relevant 'place' or 'context' is posited by each new aggregation/scale; we should not be surprised that estimates change, too (Petrović et al., 2018a). Thus, we cannot escape the MAUP by drawing more or better areal units: we must make stronger theories about what context, exposure, or place *does*.

2.iii The future of areal data

Finally, Fotheringham and Wong (1991) suggest we leave the areal unit altogether. 'Accidental' data (Arribas-Bel, 2014), our individual digital 'by-products' (Kitchin, 2013), may empower us to study people and the effect of their environs. This data is not perfect (Wyly, 2014; Ash et al., 2018), but can be useful with the right theoretical grounding (Shelton et al., 2015: 199). Indeed, bias from non-representative samples (Boeing and Waddell, 2017; Folch et al., 2018; Zhang and Zhu, 2018) or spatial disparity in coverage (Shelton et al., 2014) is increasingly well studied.

However, ‘intentional’ census data is not representative, either. Census units are drawn to facilitate data collection and comparison over time (US Census Bureau, 2007). While Krieger (2006) claims estimates from the 2000 US Census areas are ‘on par with those obtained with individual-level socioeconomic measures’ (Krieger, 2006: 358), recent work disagrees. Census aggregates’ uncertainty varies but is ignored (Spielman et al., 2014). Petrović et al. (2018b) find different probabilities of interaction in the Netherlands comparing individual responses to official aggregates. Fowler et al. (2020) also show the representativeness of 2010 US census data aggregates varies geographically. This does not *invalidate* “intentional” census data, but shows areal census data requires critical empiricism, too. Thus, the most plausible ‘solution’ to the MAUP may be to leave the modifiable areal unit behind.

III Forces shaping quantitative geographers

Beyond theory, the following social and material forces are changing geographers and their production of geographic knowledge.

I Reproducibility

Hardly anybody takes data analyses seriously. Or perhaps more accurately, hardly anyone takes anyone else’s data analyses seriously. (Leamer, 1983: 13)

Reproducibility is transforming science. Bollen et al.’s (2015) definition of reproducibility entails two separate but related concepts. The first, *reproducibility*, means we can ‘duplicate the results of a prior study using the same materials and procedures’ (p. 3). The second, *replicability*, means we can ‘duplicate the results of a prior study if the same procedures are followed but new data are collected’ (p. 4). Thus, while *reproducibility* might be attained by code and

data sharing, *replicability* is an epistemological standard of generalization which the theories discussed in previous sections must pass.

Failure to replicate has challenged accumulated knowledge across social science (Ioannidis, 2015; Collaboration, 2015). Projects like *Retraction Watch* and *PubPeer* provide a platform for community critique where bloggers publicly shame offending scholars in hopes of discouraging future malfeasance (Didier and Guaspere-Cartron, 2018: 165–6). Often, these retractions involve replication failures, but Moylan and Kowalczyk (2016) find retractions are still driven by misconduct which requires structural fixes to prevent (Munafò et al., 2018). Regardless, we should not forget: ethical research can also fail to replicate.

Whether this transfers to geography remains to be seen. Brunson (2016) outlines the challenges posed for reproducibility in geography, and Kedron et al. (2020) provide a further elaboration of the theoretical challenges to replication in geography. Further, *Retraction Watch* shows retractions in a few geography journals, but these are rare and will hopefully remain so. Some welcome ‘disciplining’ of poor practice (O’Loughlin, 2018a, 2018b), but this remains controversial.

If replication remains inhibited by the field’s theoretical issues discussed in Section II, it will remain elusive. For instance, is Church (2018) a ‘replication’ of Tobler’s Law because it proved useful in simplifying a problem? Or, is Tobler’s Law only replicated by distance decays in spatial auto-covariance functions? Regardless, replication entails a new epistemological standard that is changing other branches of science. Quantitative geography will be stronger if it embraces replication, too.

2 Inclusion

One opportunity for replication will come from replicating research across a large and diverse set of places by people with different perspectives.

This requires a more inclusive geography in both scope and composition. Our future, in terms of the people that inhabit it and the places it studies, must be broader than our past.

2.i Inclusion as self-awareness

One area where geography can become more inclusive is in its undergraduate intake. Dorling's (2019) recent reflections on his time as an admissions officer are illustrative:

I have appeared to sanction what are clearly discriminatory practices . . . I have interviewed children at the University of Bristol who were in want of a place to study Geography and, because I was told I should do so, I turned the majority away. [. . .] In reality, I thought that I had the ability to spot the potential in others – how wrong I was! (p. 4)

Indeed, throughout the 20th century, setting reasonably high academic standards gave elite American universities the ability to select students using potentially discriminatory 'non-academic' factors (Karabel, 2006: 292). Thus, exclusiveness in academia has perpetuated inequality. Instead, Crow and Dabars (2015) argue universities should be 'measured not by those whom we exclude, but rather by those whom we include and how they succeed' (p. 242).

While Crow and Dabars (2015) and Dorling (2019) write about undergraduate admissions, they have a wider point. Elite institutions make safe bets on students whose life-courses hardly improve from having been admitted. Further, these same institutions make safe hires to consolidate their leads: the best want to be at powerful programs and powerful programs want the best. At no point is there any fundamental interrogation beyond superfluous interview questions: why here, why now, why us? The answer – because this has been the top program for years – remains unspoken.

Despite our many shared beliefs and objectives, a new radical self-knowledge must

involve a recognition of the *limits of our capacity to know the potential of others*. Thus, in absence of perfect knowledge, we should be inclusive: embrace perspectives that are different from our own *even if we do not yet know the precise benefit*. Quantitative geographers have a lot to learn: we should include new and different voices in our conversations.

2.ii Moving beyond our old bounding box

We also need to extend the bounding box of our domain. There is immense opportunity for a truly *global* quantitative geography. Replication requires new and different settings in which we test our theories. And, we must not forget that "our theories" includes non-Western geographers with ideas not yet published in English. Most people who benefit from our replicable research will not inhabit the places that currently dominate our journals.

An adjacent field, city science, is instructive here. City science focuses on empirical regularities of city systems. In doing so, city scientists seek empirical regularities *across* cities and, less frequently, structural differences *between* cities (Brelsford et al., 2018; Boeing, 2019). While structural differences are important, the search for commonality *across cities* is the target of replication; cities with different histories, cultures, ages, locales, or sizes provide the points at which replication succeeds or fails.

Quantitative geography does not have a coherent subject like city science, so it has been difficult to agree on replication targets. Fortunately, replication will benefit many other parts of quantitative geography's future, so including more of the world in our science will help us across the board.

3 Focusing conversations on common tasks

We have [. . .] many 'approaches' but few arrivals. (Merton, 1949: 458)

Beyond inclusion, quantitative geography requires more direct engagement. A former quantitative geography journal editor argued that replication has not caught on as much in geography because geographers are not paying attention to one another (pers. comm., 2017). We must fix this: paying diligent attention to other scholars and developing or incorporating constructive criticism is difficult but necessary for our future.

One way to focus attention is what Donoho (2017) calls the ‘common task framework’ (CTF). In a CTF, a practical common challenge is identified and teams of scholars compete to provide the best solution. This constructive competition has been a critical driver of progress in data science and machine learning. Further, as Watts (2017) argues, the process of specifying and solving practical problems can invigorate social science, since it forces scholars to find common ground about measures, theories, and methods.

In fact, CTFs exist in some areas of quantitative geography already. Beyond clear examples in spatial statistics (Heaton et al., 2019), Steinitz’s (2012) ‘geodesign’, an ambitious repurposing and extension of GIScience (Goodchild, 2010; Batty, 2013), provides guidelines for successful CTFs in planning (IGC, 2020). For example, the 2020 International Geodesign Collaboration (IGC, 2020) gathers teams of scholars to build solutions to urban planning challenges using common empirical measures (e.g. Millennium Sustainable Development Goals) and cartographic/written styles. While this standardization is not without critique (Wilson, 2015; Elwood, 2006), CTFs gain prominence because they are accessible to many (*contra* Cressie and Wikle, 2017), are challenging, and have societal value if solved. Thus, a better quantitative geography will involve more attentive conversations, including constructive critique about solutions to common problems.

4 Disabling technologies

The emergence of geographic information science has long been intertwined with the maturation of systems for processing geographical data (Goodchild, 1992). One foundational insecurity of GIScience is that our notation, either intrinsically or through its computational implementation, might affect the questions we think are valuable to answer. Whereas Iverson (1980) argued that computational notation can be more useful than mathematical notation, Gahegan (1999) drew attention to the constraints GIS placed on geographic thinking. He claimed ‘generic solutions’ to data processing problems forced us to ‘adopt impoverished representational and analysis capabilities . . . in exchange for ditching the Fortran, getting some sleep and producing much prettier output’. Fundamentally, Gahegan’s (1999) concern is about the limits our computational tools place on our thinking; our ‘technical debt’, the accumulated technological and social costs associated with past engineering decisions, may leave us penniless, with no new theory in the bank.

This concern is still with us. Programming in GIScience instruction is on the rise (Etherington, 2016; Bowlick et al., 2017; Arribas-Bel, 2019), underwritten by high-quality spatial analysis libraries (Rey and Anselin, 2007; Bivand et al., 2011) and pedagogy treating ‘code as text’ (Rey, 2009, 2018). This has resulted in community-led scientific infrastructure (Wolf et al., 2019) where international teams contribute to a shared body of implemented knowledge.

As data gets larger and techniques more complicated, specialized high-performance computing frameworks have become more common in cutting-edge geographical analysis. ‘Distributed’ computing libraries, such as spark (Zaharia et al., 2016), can spread computational load across many computers. Alternatively, ‘tensor’ computing libraries, such as tensorflow (Abadi et al., 2016) or PyTorch (Paszke et al., 2019),

automatically optimize large chains of computations required by machine learning. This is not to say that *all* quantitative geography requires substantial computing, but *many* new methods will be inaccessible to geographers without these frameworks. Unfortunately, these are next-generation disabling technologies in two senses.

4.i Platform capitalism

First, these libraries are large corporate-led open source projects. Indeed, the two main tensor computing frameworks are dominated by Google and Facebook. This complicates the political economy of community-led scientific infrastructure, since large companies realized:

Sharing, rather than building proprietary code, turned out to be cheaper, easier, and more efficient. This increased demand puts additional strain on those who maintain this infrastructure, yet because these communities are not highly visible, the rest of the world has been slow to notice. (Eghbal, 2016: 9)

Further, companies also support these libraries as part of a broader turn towards ‘platform capitalism’ (Srniczek, 2017). These companies seek ‘mindshare’, a neologism describing psychological influence over scientists, affecting how science is done on a day-to-day basis and how new science is planned. This inhibits quantitative geographers from full control over the implementation of quantitative geography. To resist this, a successful quantitative geography must reinforce community-led pedagogically-oriented scientific infrastructure.

4.ii Constraining the possible analyses

In addition, these packages are usually not designed for geographical applications. While these frameworks make it easy to implement complex neural networks, they also enforce

representations that make sense to their designers. Since these designers are optimizing for specialized industrial data science and machine learning applications, it can be challenging to build scientifically-useful geographical models.

To illustrate, Singleton and Arribas-Bel (2019) make a case for a new ‘geographic data science’ integrating geography and data science. The extent of integration may take a few forms. First, *commodity data science* (Singleton, pers. comm., 2020; Maskell, 2019) uses standard algorithms on data ignoring geography. However, the results might be visualized in a geographical way. Second, *data science enhanced with geography* enriches data with geographical information and analyses it using standard algorithms. Third, *explicitly geographic data science* harnesses the geographical structure of data to give novel insights.

Explicitly geographic methods are usually not supported or are difficult to optimize in common high-performance computing frameworks. The second ‘enhanced’ form exists at the frontier of geography (de Sabbata and Liu, 2019; Zhu et al., 2020), but is not routine. The third form also requires new foundational work. Thus, a successful quantitative geography must build its own tools or, better still, commandeer frameworks to build a fairer community-owned computational infrastructure.

5 Causality

Singleton and Arribas-Bel’s (2019) geographic data science is visionary. However, the integration they see is not sufficient to define a new domain beyond geography itself. Replacing data science for something else only changes what gets integrated with geography: all quantitative geographic methods exist in a similar hierarchy of integration. This also applies to the burgeoning field of causal inference.

While ‘causal geographies’ are still elusive, geography increasingly *enhances* standard

causal inference methods. One clear case, *regression discontinuity designs* (RDDs), estimate the causal impact of an intervention applied according to a fixed, exogenous threshold (Imbens and Lemieux, 2008). This threshold is the ‘discontinuity’ separating two groups that are otherwise fundamentally similar.

For geographical problems, *boundaries* provide this discontinuity: adjacent areas may be demographically similar, but laws change abruptly between jurisdictions. The causal effect of a new law can thus be estimated from boundary discontinuities. Despite controversy (Chen et al., 2013; Pope and Dockery, 2013), geographic RDDs (Keele and Titiunik, 2016) saw early use in the analysis of government policy (Holmes, 1998) and have recently been used to study financial risk (Goetz et al., 2016), crime and policing (MacDonald et al., 2016; Twinam, 2017), electoral turnout (Keele and Titiunik, 2018), and school quality (Gibbons et al., 2013). Because boundaries are intrinsic to administrative/areal data, geo-RDDs remain a causal inference method of choice.

However, geo-RDDs in quantitative political geography exhibit what O’Loughlin (2018b) calls ‘political geometry’. Among other methods, geo-RDDs ‘reject the possibility of contextual effects that complicate the usual socio-demographic or ideological predictors of political behavior’ (O’Loughlin, 2018a, 2018b). Beyond political geography, this means the effect of ‘place’ is modelled using a ‘distance to boundary’ measure. This leaves no room for substantive place-based differences: one is either ‘near’ or ‘far’, ‘in’ or ‘out’. Thus, when results are sensitive to changes in distance metrics, something more than the discontinuity must be at play.

We must approach this carefully. The ‘Cartesian coordinate approach’ setting up the title fight between ‘The Good, the Bad, and the Ugly’ in O’Loughlin (2018b) is only a ‘general

orientation towards data’ (Merton, 1949). But, changing a distance metric is a theoretical act with empirical effects. If the distances core to our ‘Cartesian’ approaches are only fungible in theory, then better theory is needed. Indeed, ‘testing theories means correctly estimating the coefficients on specific causal variables’ (Gibbons and Overman, 2012: 186). There is no reason to believe that the various theories implied by different notions of ‘near’ should yield identical causal coefficients. As such, we must redouble our efforts to engage with the theoretical implications of causality in geography. Indeed, for us to ever conduct replicable research, geographers must explicitly define specific concepts, theories, causes, and processes to continue to integrate successfully with other domains.

IV Conclusion

In this report, we outlined theoretical and social challenges that quantitative geography faces. Broadly, if we are to have a future, we must strive to be

- **replicable** in our design
- **inclusive** in our self-concept
- **specific** in our theory
- **open** in our execution

This may manifest in a few ways.

For our theory, we may broaden ideas considered here or develop entirely new heuristics. With a better Tobler’s Law, we might help other domains simplify difficult analytical problems – after all, geographical data is pervasive. With a better theoretical understanding of the MAUP, we become more explicit about how changes in aggregation change the theory of context, place, or exposure within ethically-sourced secure but accessible individual data.

Practically, challenges to open and reproducible quantitative geography will be overcome by strengthening our community. Resisting pressures from large-scale single-party control

of computational frameworks will not be easy. Access to open source scientific tools is improving, and the maintenance of this critical digital infrastructure may still return to the commons. Further, successful campaigns to build a more inclusive scientific community need quantitative geographers to play their part in making our own community more inclusive in focus and composition.

Beyond community, replication will still be challenging. We will continue to find success by taking causality seriously, building specific testable theories amenable to replication. With better geographic data, we will need better geographic theory. What does this new geography ‘do’, beyond acting as a proxy for something else? Our aspirations about ‘contextual effects’ should motivate us to seek replicable designs that explicitly theorize what place, context, or exposure do in geographical processes. Alternatively, what assumptions about place, context, or exposure are already embedded in our current notions of ‘distance’ or ‘region’?

In some cases, the effect of geography (be it as Tobler’s ‘space’ or Openshaw’s ‘place’) may boil down to a pure geometric relationship – in or out, near or far. If so, we should seek to replicate the result, and examine where replication fails and why. If not, this simplicity should not be disappointing. It is the erosion of complexity – slowly and progressively by accumulated knowledge – that is the point.


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

Levi John Wolf  <https://orcid.org/0000-0003-0274-599X>

Ron Johnston  <https://orcid.org/0000-0003-3031-8573>

Emmanouil Tranos  <https://orcid.org/0000-0002-9620-6542>

References

- Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J and Devin M (2016) TensorFlow: A system for large-scale machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. Available at: https://www.usenix.org/sites/default/files/osdi16_full_proceedings.pdf (accessed 28 April 2020).
- Arribas-Bel D (2014) Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography* 49(May): 45–53. DOI: 10.1016/j.apgeog.2013.09.012.
- Arribas-Bel D (2019) A course on geographic data science. *Journal of Open Source Education* 2(14): 42. DOI: 10.21105/jose.00042.
- Ash J, Kitchin R and Leszczynski A (2018) Digital turn, digital geographies? *Progress in Human Geography* 42(1): 25–43. DOI: 10.1177/0309132516664800.
- Batty M (2013) Defining Geodesign (= GIS + Design?). *Environment and Planning B: Planning and Design* 40(1): 1–2. DOI: 10.1068/b4001ed.
- Bivand R, Anselin L, Berke O, Bernat A, Carvalho M, Chun Y and Dormann CF (2011) *Spdep: Spatial Dependence: Weighting Schemes, Statistics and Models* [R package version 0.5-31]. Available at: <http://CRAN.R-project.org/package=spdep> (accessed 28 April 2020).
- Boeing G (2019) Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science* 4(1): 67. DOI: 10.1007/s41109-019-0189-1.
- Boeing G and Waddell P (2017) New insights into rental housing markets across the United States: Web scraping and analyzing Craigslist rental listings. *Journal of Planning Education and Research* 37(4): 457–476. DOI: 10.1177/0739456X16664789.
- Bollen K, Cacioppo J, Kaplan R, Krosnick J and Olds J (2015) *Social, Behavioral, and Economic Sciences: Perspectives on Robust and Reliable Science*. Arlington, VA: National Science Foundation.
- Bowlick FJ, Goldberg DW and Bednarz SW (2017) Computer science and programming courses in geography departments in the United States. *The Professional*

- Geographer* 69(1): 138–150. DOI: 10.1080/00330124.2016.1184984.
- Bradley JR, Wikle CK and Holan SH (2016) Regionalization of multiscale spatial processes by using a criterion for spatial aggregation error. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79(3): 815–832.
- Brelsford C, Martin T, Hand J and Bettencourt LMA (2018) Toward cities without slums: Topology and the spatial evolution of neighborhoods. *Science Advances* 4(8). DOI: 10.1126/sciadv.aar4644.
- Broido AD and Clauset A (2019) Scale-free networks are rare. *Nature Communications* 10(1): 1–10. DOI: 10.1038/s41467-019-08746-5.
- Brunsdon C (2016) Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography* 40(5): 687–696. DOI: 10.1177/0309132515599625.
- Cairncross F (2001) *The Death of Distance: How the Communications Revolution Is Changing Our Lives (new edition)*. Boston, MA: Harvard Business School Press.
- Chen Y, Ebenstein A, Greenstone M and Li H (2013) Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences* 110(32): 12936–12941. DOI: 10.1073/pnas.1300018110.
- Church RL (2018) Tobler's Law and spatial optimization: Why Bakersfield? *International Regional Science Review* 41(3): 287–310.
- Collaboration, Open Science (2015) Estimating the reproducibility of psychological science. *Science* 349(6251). DOI: 10.1126/science.aac4716.
- Cressie N and Wikle C (2017) *Common Task Framework for Spatial Prediction*. Available at: <https://hpc.niasra.uow.edu.au/ctf/> (accessed 28 April 2020).
- Crow MM and Dabars WB (2015) *Designing the New American University*. Baltimore, MD: Johns Hopkins University Press.
- de Sabbata S and Liu P (2019) Deep learning geodemographics with autoencoders and geographic convolution. *Proceedings of the 22nd AGILE conference on Geographic Information Science*, Limassol, Greece.
- Didier E and Guaspere-Cartron C (2018) The new watchdogs' vision of science: A roundtable with Ivan Oransky (*Retraction Watch*) and Brandon Stell (*PubPeer*). *Social Studies of Science* 48(1): 165–67. DOI: 10.1177/0306312718756202.
- Donoho D (2017) 50 years of data science. *Journal of Computational and Graphical Statistics* 26(4): 745–766. DOI: 10.1080/10618600.2017.1384734.
- Dorling D (2019) Kindness: A new kind of rigour for British geographers. *Emotion, Space and Society* 33(November): 100630. DOI: 10.1016/j.emospa.2019.100630.
- Duque JC, Laniado H and Polo A (2018) S-Maup: Statistical test to measure the sensitivity to the Modifiable Areal Unit Problem. *PLOS ONE* 13(11): e0207377. DOI: 10.1371/journal.pone.0207377.
- Eghbal N (2016) *Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure*. Washington, DC: Ford Foundation.
- Elwood S (2006) Critical issues in participatory GIS: Deconstructions, reconstructions, and new research directions. *Transactions in GIS* 10(5): 693–708.
- Etherington T (2016) Teaching introductory GIS programming to geographers using an open source Python approach. *Journal of Geography in Higher Education* 40(1): 117–130.
- Folch DC, Spielman SE and Manduca R (2018) Fast food data: Where user-generated content works and where it does not. *Geographical Analysis* 50(2): 125–40. DOI: 10.1111/gean.12149.
- Fotheringham AS and Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23(7): 1025–1044.
- Fotheringham AS, Brunsdon C and Charlton M (2000) *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: SAGE.
- Fowler CS, Frey N, Folch DC, Nagle N and Spielman S (2020) Who are the people in my neighborhood?: The 'contextual fallacy' of measuring individual context with census geographies. *Geographical Analysis*. DOI: 10.1111/gean.12192.
- Gahegan M (1999) Guest editorial: What is geocomputation? *Transactions in GIS* 3: 203–206.
- Gibbons S and Overman HG (2012) Mostly pointless spatial econometrics? *Journal of Regional Science* 52(2): 172–191. DOI: 10.1111/j.1467-9787.2012.00760.x.
- Gibbons S, Machin S and Silva O (2013) Valuing school quality using boundary discontinuities. *Journal of Urban Economics* 75(May): 15–28. DOI: 10.1016/j.jue.2012.11.001.
- Goetz MR, Laeven L and Levine R (2016) Does the geographic expansion of banks reduce risk? *Journal of*

- Financial Economics* 120(2): 346–362. DOI: 10.1016/j.jfineco.2016.01.020.
- Golledge RG (2002) The nature of geographic knowledge. *Annals of the Association of American Geographers* 92(1): 1–14. DOI: 10.1111/1467-8306.00276.
- Goodchild MF (1992) Geographical information science. *International Journal of Geographical Information Science* 6(1): 31–45.
- Goodchild MF (2004) The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94(2): 300–303.
- Goodchild MF (2010) Towards geodesign: Repurposing cartography and GIS? *Cartographic Perspectives* 66(June): 7–22. DOI: 10.14714/CP66.93.
- Harvey D (1969) *Explanation in Geography*. London: Hodder & Stoughton Educational.
- Heaton MJ, Datta A, Finley AO, Furrer R, Guinness J, Guhaniyogi R and Gerber F (2019) A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics* 24(3): 398–425. DOI: 10.1007/s13253-018-00348-w.
- Hecht B and Moxley E (2009) Terabytes of Tobler: Evaluating the First Law in a massive, domain-neutral representation of world knowledge. In: Hornsby KS, Claramunt C, Denis M and Ligozat G (eds) *Spatial Information Theory*. Berlin: Springer, 88–105. DOI: 10.1007/978-3-642-03832-7_6.
- Holmes TJ (1998) The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy* 106(4): 667–705. DOI: 10.1086/250026.
- IGC (2020) Projects workflow: How will IGC studies be carried out? *International Geodesign Collaboration*. Available at: <https://www.igc-geodesign.org/project-workflow> (accessed 28 April 2020).
- Imbens GW and Lemieux T (2008) Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2): 615–635. DOI: 10.1016/j.jeconom.2007.05.001.
- Ioannidis JPA (2015) Why most published research findings are false. *PLOS Medicine* 2(8): e124. DOI: 10.1371/journal.pmed.0020124.
- Isard W (1956) Regional science, the concept of region, and regional structure. *Papers in Regional Science* 2(1): 13–26.
- Iverson KE (1980) Notation as a tool of thought. *Communications of the ACM*. DOI: 10.1145/358896.358899.
- Johnston R and Sidaway JD (2015) *Geography and Geographers: Anglo-American Human Geography since 1945*. New York: Routledge.
- Johnston R, Jones K, Burgess S, Propper C, Sarker R and Bolster A (2004) Scale, factor analyses, and neighborhood effects. *Geographical Analysis* 36(4): 350–368. DOI: 10.1111/j.1538-4632.2004.tb01141.x.
- Jones KT (1998) Scale as epistemology. *Political Geography* 17(1): 25–28. DOI: 10.1016/S0962-6298(97)00049-8.
- Karabel J (2006) *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. New York: Houghton Mifflin Harcourt.
- Kedron P, Frazier AE, Trgovac AB, Nelson T and Fotheringham AS (2020) Reproducibility and replicability in geographical analysis. *Geographical Analysis*. DOI: 10.1111/gean.12221.
- Keele L and Titiunik R (2016) Natural experiments based on geography. *Political Science Research and Methods* 4(1): 65–95. DOI: 10.1017/psrm.2015.4.
- Keele L and Titiunik R (2018) Geographic natural experiments with interference: The effect of all-mail voting on turnout in Colorado. *CESifo Economic Studies* 64(2): 127–149. DOI: 10.1093/cesifo/ify004.
- King G (1996) Why context should not count. *Political Geography* 15(2): 159–164.
- Kitchin R (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267. DOI: 10.1177/2043820613513388.
- Krieger N (2006) A century of census tracts: Health & the Body Politic (1906–2006). *Journal of Urban Health* 83(3): 355–361.
- Kwan M-P (2018) The limits of the neighborhood effect: Contextual uncertainties in geographic, environmental health, and social science research. *Annals of the American Association of Geographers*. DOI: 10.1080/24694452.2018.1453777.
- Leamer EE (1983) Let's take the con out of econometrics. *The American Economic Review* 73(1): 31–43.
- LeSage JP and Pace RK (2014) The biggest myth in spatial econometrics. *Econometrics* 2: 217–249.
- MacDonald JM, Klick J and Grunwald B (2016) The effect of private police on crime: Evidence from a geographic regression discontinuity design. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179(3): 831–846. DOI: 10.1111/rssa.12142.
- Maskell S (2019) *EPSRC Centre for Doctoral Training in Distributed Algorithms: The what, how and where of*

- next-generation data science. UKRI Grant Application EP-S023445 -1*. Liverpool: University of Liverpool.
- Merton RK (1949) On sociological theories of the middle range. In: Merton RK (ed.) *Social Theory and Social Structure*. New York: Simon & Schuster, 39–53.
- Miller HJ (2004) Tobler's First Law and spatial analysis. *Annals of the Association of American Geographers* 94(2): 284–289.
- Moylan EC and Kowalczyk MK (2016) Why articles are retracted: A retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open* 6(11). DOI: 10.1136/bmjopen-2016-012047.
- Munafò MR, Hollands GJ and Marteau TM (2018) Open science prevents mindless science. *BMJ* 363(October). DOI: 10.1136/bmj.k4309.
- Nelson GD (2020) Communities, complexity, and the 'conchoration': Network analysis and the ontology of geographic units. *Tijdschrift Voor Economische En Sociale Geografie*. DOI: 10.1111/tesg.12400.
- O'Loughlin J (2018a) More drilling, more transparency and less skipping: A reply to the commentators. *Political Geography* 65(July): 159–160. DOI: 10.1016/j.polgeo.2018.05.005.
- O'Loughlin J (2018b) Thirty-five years of political geography and political geography: The good, the bad and the ugly. *Political Geography* 65(July): 143–151. DOI: 10.1016/j.polgeo.2018.05.004.
- Openshaw S (1984) *The Modifiable Areal Unit Problem*. Norwich: Geo Abstracts, University of East Anglia.
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G and Killeen T (2019) PyTorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E and Garnett R (eds) *Advances in Neural Information Processing Systems* 32. Red Hook, NY: Curran Associates, Inc., 8026–8037.
- Petrović A, Manley D and Van Ham M (2018a) Freedom from the tyranny of neighbourhood: Rethinking socio-spatial context effects. *Progress in Human Geography*. DOI: 10.1177/0309132519868767.
- Petrović A, Van Ham M and Manley D (2018b) Multi-scale measures of population: Within- and between-city variation in exposure to the sociospatial context. *Annals of the American Association of Geographers* 108(4): 1057–1074. DOI: 10.1080/24694452.2017.1411245.
- Pope CA and Dockery DW (2013) Air pollution and life expectancy in China and beyond. *Proceedings of the National Academy of Sciences* 110(32): 12861–12862. DOI: 10.1073/pnas.1310925110.
- Rey SJ (2009) Show me the code: Spatial analysis and open source. *Journal of Geographical Systems* 11(2): 191–207. DOI: 10.1007/s10109-009-0086-8.
- Rey SJ (2018) Code as text: Open source lessons for geospatial research and education. In: Thill J-C and Dragicovic S (eds) *GeoComputational Analysis and Modeling of Regional Systems*. Cham: Springer, 7–21. DOI: 10.1007/978-3-319-59511-5_2.
- Rey SJ and Anselin L (2007) PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies* 37(1): 5–27.
- Shelton T, Poorthuis A, Graham M and Zook M (2014) Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. *Geoforum* 52: 167–179. DOI: 10.1016/j.geoforum.2014.01.006.
- Shelton T, Poorthuis A and Zook M (2015) Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning (Special Issue: Critical Approaches to Landscape Visualization)* 142: 198–211. DOI: 10.1016/j.landurbplan.2015.02.020.
- Singleton A and Arribas-Bel D (2019) Geographic data science. *Geographical Analysis*. DOI: 10.1111/gean.12194.
- Singleton A and Spielman SE (2014) The past, present, and future of geodemographic research in the United States and United Kingdom. *Professional Geographer* 66 (4): 558–57.
- Smith JM (2004) Unlawful relations and verbal inflation. *Annals of the Association of American Geographers* 94(2): 294–299. DOI: 10.1111/j.1467-8306.2004.09402007.x.
- Spielman SE and Folch DC (2015) Reducing uncertainty in the American Community Survey through data-driven regionalization. *PLoS One* 10(2): e0115626.
- Spielman SE, Folch DC and Nagle N (2014) Patterns and causes of uncertainty in the American Community Survey. *Applied Geography (Sevenoaks, England)* 46 (January): 147–157. <https://doi.org/10.1016/j.apgeog.2013.11.002>.
- Srnicek N (2017) *Platform Capitalism*. John Wiley & Sons.
- Steinitz C (2012) *A Framework for Geodesign: Changing Geography by Design*. Redlands, CA: ESRI Press.

- Tam Cho WK (2018) Algorithms can foster a more democratic society. *Nature* 558(7711): 487. DOI: 10.1038/d41586-018-05498-y.
- Tam Cho WK and Liu YY (2018) Sampling from complicated and unknown distributions: Monte Carlo and Markov Chain Monte Carlo methods for redistricting. *Physica A: Statistical Mechanics and Its Applications* 506 (September): 170–178. DOI: 10.1016/j.physa.2018.03.096.
- Tapp K (2019) Measuring political gerrymandering. *The American Mathematical Monthly* 126(7): 593–609. DOI: 10.1080/00029890.2019.1609324.
- Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(June): 234. DOI: 10.2307/143141.
- Tobler WR (1989) Frame independent spatial analysis. *Accuracy of Spatial Databases*, 115–22.
- Tranos E and Nijkamp P (2013) The death of distance revisited: Cyber-place, physical and relational proximities. *Journal of Regional Science* 53(5): 855–873. DOI: 10.1111/jors.12021.
- Tukey J (1962) The future of data analysis. *The Annals of Mathematical Statistics* 33(1): 1–67.
- Twinam T (2017) Danger zone: Land use and the geography of neighborhood crime. *Journal of Urban Economics* 100(July): 104–119. DOI: 10.1016/j.jue.2017.05.006.
- US Census Bureau (2007) *Census Tract Program for the 2010 Census-Proposed Criteria*. Washington, DC: National Archives.
- Watts DJ (2017) Should social science be more solution-oriented? *Nature Human Behaviour* 1(1): 1–5. DOI: 10.1038/s41562-016-0015.
- Wilson MW (2015) On the criticality of mapping practices: Geodesign as critical GIS? *Landscape and Urban Planning (Special Issue: Critical Approaches to Landscape Visualization)* 142(October): 226–234. DOI: 10.1016/j.landurbplan.2013.12.017.
- Wolf LJ, Rey SL and Oshan TM (2019) Open code is not enough: Towards a replicable future for geographic data science. In: *Third National Science Foundation Workshop on Conceptualizing a Geospatial Software Institute*. SocArXiv. DOI: 10.31235/osf.io/3hbnt.
- Wyly E (2014) The new quantitative revolution. *Dialogues in Human Geography* 4(1): 26–38. DOI: 10.1177/2043820614525732.
- Zaharia M, Xin RS, Wendell P, Das T, Armbrust M, Ankur D and Xiangrui M (2016) Apache Spark: A unified engine for big data processing. *Communications of the ACM* 59(11): 56–65. DOI: 10.1145/2934664.
- Zhang G and Zhu A-X (2018) The representativeness and spatial bias of volunteered geographic information: A review. *Annals of GIS* 24(3): 151–162. DOI: 10.1080/19475683.2018.1501607.
- Zhu D, Zhang F, Wang S, Wang Y, Cheng X, Huang Z and Liu Y (2020) Understanding place characteristics in geographic contexts through graph convolutional neural networks. *Annals of the American Association of Geographers* 110(2): 408–420. DOI: 10.1080/24694452.2019.1694403.