Reproducibility and Replicability in Geographical Analysis

Peter Kedron, Amy Frazier, Andrew Trgovac, Trisalyn Nelson, Stewart Fotheringham

Preprint of Kedron, P., Frazier, A.E., Trgovac, A., Nelson, T., Fotheringham, S. (2021) Reproducibility and replicability in geographical analysis. *Geographical Analysis* 53: 135-147

Abstract. The scientific method is predicated on the assumption that research designs and results can be reproduced and replicated. However, recent findings in some disciplines suggest that many studies fail to reach this standard, moving issues surrounding reproducibility and replicability of forward in the research agenda of those fields. While the topic has yet to become a point of controversy in geography, the intricacies of geographic phenomena and spatial data analysis make the field vulnerable to criticism. This commentary discusses how uncertainties related to the conception, measurement, analysis, and communication of geographic analyses contribute to difficulties in the reproduction and replication of geographic research. Investigating how these uncertainties collectively impact the reproducibility and replicability of spatial data analyses should be a critical focus of future *Geographical Analysis* research. A call to action for geographers to improve the reproducibility and replicability of their work and specific recommendations on how *Geographical Analysis* might facilitate this process conclude the commentary.

Introduction

The reproducibility and replicability (R&R) of scientific results has been receiving widespread attention in many fields of study because the ability to independently verify results is the fundamental, self-correcting mechanism of the scientific method. Whether or not research results can be independently reproduced or replicated is a question that has moved forward on the research agendas of many fields (Ioannidis 2005, Camerer et al. 2016, Baker 2017, Ioannidis et al. 2017). Given the important role R&R plays in scientific progress, it is somewhat surprising that the topic has received limited attention and formal study in geography and geographical analysis. While a series of recent studies and workshops has sparked a discussion of R&R in geography by assessing whether specific geographical analyses can be reproduced (SPARC Workshop 2019, Ostermann and Granell 2017, Nust et al. 2018, Konkol et al. 2019) and how the discipline might move toward broader adoption of R&R standards (Brunsdon and Singleton 2015, Brunsdon 2016), far more work is needed. Basic questions remain unanswered such as to what extent is R&R possible in geographical analysis given the roles spatial context and observation play in the discipline? In which areas of geographic research are failures to reproduce or replicate most likely or most severe? And, how are expectations about R&R and failures to reproduce or replicate related to the fundamental uncertainties inherent in any spatial analysis? Looking to the future of geography, spatial science, and geographical analysis, we argue that R&R must move forward on the discipline's research agenda. Improving our ability to explain geographic phenomena rests on our ability to understand, verify, apply, and extend the inferences of prior research.

Conflicting definitions of 'reproducible' and 'replicable' have emerged in the literature as disciplinary interest in the subject has expanded (Barba 2018, Plesser 2018). While debate about consistent terminology continues, the most widely adopted definitions follow a convention linked to Claerbout and Karrenbach (1992), further formalized by Donoho et al. (2009) and Peng (2011). In 2015, The National Science Foundation embraced these definitions and formally defined the terminology as follows (Bollen et al. 2015):

> ***Reproducibility*** *refers to the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator.*
>
> ***Replicability*** *refers to the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected*

Stated more succinctly, research is *reproducible* when an independent researcher can use the same data and the same methods to produce the same results (same data + same methods = same results). Research is *replicable* when new data are collected, but the same (or very similar) methods produce the same (or very similar) results (new data + same/similar methods = same/similar results). The precise definition of same and similar remains ill-defined in many fields. Perhaps the best current discussion of this topic can be found in Goodman et al. (2016), Peng (2011), and Stodden (2015), which respectively link R&R to specific objectives (obtaining the same results or inferences), causes (identifying why results differ), and degrees (placing results along a spectrum of R&R). Geography is an inherently interdisciplinary field that leverages several scientific paradigms to make statements about spatial generating processes. Therefore, it is reasonable for geographers to draw from each of these three interpretations. Here, we use the acronym R&R to refer generally to the issues at hand, and we use the terms 'reproducibility' and 'replication' according to the NSF definitions above. Our primary interest though is in the implications of R&R for inference. Ultimately, developing reproducible and replicable research practices in geography while understanding the limits of R&R will allow us to create inferences that are more robust to changes in data or context and lead to the development of internally consistent and generalizable theory – better explanation of geographic phenomena.

With 50 years of involvement in the field, *Geographical Analysis,* is well-positioned to lead the study of reproducibility and replicability issues in geography as well as the development of standards and practices that improve our ability to explain geographic phenomena. As an early contribution to this stream of research, this paper provides a commentary on how the uncertainties of geographical analysis relate to the reproducibility and replicability of results and inference. The focus of our discussion is R&R, but we use uncertainty as a means of placing the causes of failures to reproduce or replicate research at specific points within the stages of the scientific method. This approach allows us to present a framework for future R&R research that will allow researchers to place their contributions within the scientific research process. While a comprehensive assessment of these issues is beyond the scope of any single work, we highlight critical features of geographical analysis likely to affect R&R, and discuss how geographers can contribute to this stream or research.

Reproducibility, Replicability, and Uncertainties in Geographical Analysis

Uncertainty is present in geographical analyses because it is not possible to completely represent the complexity of the real world. In this regard, geographical analysis shares a challenge faced by any empirical science. Uncertainty about how best to conceptualize, measure, analyze, or communicate research expands the number of choices and decisions researchers must make during the course of an investigation. When we seek to replicate scientific findings, the presence of uncertainties in prior studies can lead different researchers to make alternative choices and decisions, even when examining the same process in the same spatial context. Within high-dimensional datasets, the sheer number of variables and the presence of natural variation also expands the number of reasonable approaches to analyzing the same or different data (Munafo et al. 2017).

Facing uncertainty and natural variation in collected data, geographers often “try and select” (sensu Gelman and Loken 2013; de Groot 1956) different decisions en route to a final analysis, result, and inference. During that process, it can be difficult for researchers to recognize, and later communicate, all the choices they made, and it remains common practice to only report the single result and narrow set of choices that generated a specific outcome. Termed ‘researcher degrees of freedom’ (Simmons et al. 2011), this collective set of often unreported choices, and associated unreported alternative analyses, limits our ability to consider results and inferences in the context of their complete design and development. While the results of such a study may be reproducible, the inferences we draw from the reproduction may differ depending on how a researcher views the conceptualizations, measurements, and analyses conducted. Replication of results or inferences also becomes challenging because a failure to communicate how phenomena are conceptualized and measured limits our ability to collect comparable datasets and repeat analyses.

Further consideration of the relationship between uncertainty and R&R is warranted in geographical analysis due to the central role spatial context and spatial relationships play in the emergence of geographic phenomena. Because many geographical analyses rely on observational data drawn from uncontrolled settings, replicating geographic research requires a careful consideration of how uncertainties in the original research design were addressed and to what extent the underlying spatial generating processes should be expected to operate in the same way given a change in context that often accompanies a change in data. Dating back to at least the Hartshorne-Schaefer debate, the geographic literature has long examined whether, and in which instances, we might expect geographic processes to remain stable, thereby allowing the study of those processes to be replicable (Hartshorne 1939, Schaefer 1953, Hartshorne 1955). At the same time, the observation that geographic data tend to exhibit spatial dependencies brings into question how and to what extent replications should account for spatial relationships in ways that are consistent with prior work, while also, as far as possible, matching relationships that exist in reality. For example, it is possible to image two datasets collected independently from the same location at two different times, and for the spatial structure of a key relationship among the variables to have changed over time between the data collection periods. If the analysis of the first dataset correctly specifies the spatial dependence in the first period, should that specification then be used to set up a replication study that uses the dataset from the second time period? If the same spatial weighting method is used, the replication may fail due to a misspecification of the spatial dependence in the second period. However, it may also be the case that if the spatial weights in the second period are adjusted to match the true spatial relationship of that period, estimates might align and produce a successful replication. Determining and communicating how large a difference in the structure of spatial dependence can be between studies will be central to identifying whether a geographical study produces a replication of a prior result. Moreover, what is identified as different enough to qualify as non-replication will likely note be the same for different spatial processes and geographic contexts.

Defining the geographic contexts in which a spatial generating process produces a consistent effect or spatial pattern remains a central goal of geographic research. Making that determination depends on our ability to trust replications and reproductions of prior geographical analyses, which in turn requires a clear understanding of how uncertainties in the design and analysis shape geographic research. We therefore use uncertainty to relate characteristics of geographical analysis to R&R and place those challenges within the context of the scientific method. The study of uncertainty is well-established in geography (Goodchild 2018, Griffith 2018). We build on this tradition by integrating the summative assessment of Longley et al. (2015) with the elaboration of threats to scientific R&R presented by Munafo et al. (2017) (Fig. 1). Specifically, we examine how uncertainties in (1) conceptualization, (2) measurement, (3) analysis, and (4) communication complicate geographical analysis in ways that affect our ability to reproduce and replicate prior findings. The interior black circle in Figure 1 represents the scientific research cycle from generating a hypothesis through communication of results. Around the outside are the four types of uncertainty positioned along the portions of the scientific method they most impact. The gradual increase in thickness of the spiral represents how uncertainty and natural variation expand the number of choices a researcher is required to make, and the widening set of reasonable analyses he or she might undertake. In the following sub-sections, we use each of the four uncertainty filters as entry points to discuss the characteristics of geographical analysis that complicate the reproduction and replication of geographic research.

<< Insert Figure 1 Here >>

Conceptual uncertainty

Producing the same results and inferences during the reproduction or replication of a prior study will depend in part on whether researchers share a conceptualization of the underlying phenomena. However, matching the conception of geographic phenomena across studies can be difficult because geographic phenomena often have no natural unit of analysis, which makes their definition subject to the vagueness and ambiguity introduced by individual perceptions. These conceptual uncertainties are central to R&R because they ultimately influence how we measure, analyze, and interpret geographic phenomena.

Conceptual uncertainty typically arises early in the scientific process during hypothesis generation and study design (Fig. 1) when researchers are making decisions about what aspects of reality will be represented, which data models will be used to represent phenomena, the spatial extent of the study area, and choices surrounding scale (e.g., resolution). Conceptual uncertainty does not necessarily impact reproducibility, but it is likely to affect replication if the operationalization and communication of these decisions is not clear as that can lead to different understandings of what is being studied. If the formalization and operationalization of the concept under examination is not communicated properly by the original researcher, another researcher attempting to replicate the study may collect data that represents an entirely different aspect of a geographic phenomenon. This misalignment between concept and data shifts the purpose of the replication study from the confirmation and expansion of prior work to the exploration of new aspects of a geographic process.

Our ability to conduct replication studies of geographic phenomena can be improved by further developing ontologies that formalize geographic concepts and reduce conceptual uncertainty (Smith and Mark 2001). A classic example in geographical ontology is whether or not mountains exist (Smith and Mark 2003). Many would answer yes, but determining where a mountain ends and the valley begins is a non-trivial task. If poorly communicated, different conceptualizations could lead to different data, results, and inferences during a replication. Geographical analysis had a brief and productive history of ontological investigation around the turn of the 21st century, particularly with respect to GIS. However, activity in this domain has slowed, and ontological training is rarely part of GIScience curriculum and research. A recent search of the articles in *Geographical Analysis* revealed only 15 that mentioned ontology(ies). Encouraging ontological investigations and formalization of geographic concepts, objects, features, and relations is one way *Geographical Analysis* can positively improve our capacity to replicate geographic research.

The development of geographic ontologies is inextricably tied to scale. When a researcher collects original data, operationalization of a formal ontology rests on the practical delineation of individual units of analysis and study area extent. As a practical matter, geographers often use data sets produced or generated by others. When this occurs, a researcher may inherit a conceptualization *from* the data, which has been delivered at a predetermined scale. Building on the example above, even with a formal ontology of a mountain, a classification algorithm trained to delineate mountains from 100 m input data would produce different results compared to an algorithm trained from 5 m data.

At the same time, it is difficult to incorporate scale into geographic ontologies because many geographic phenomena have no single, characteristic scale. Complex systems, such as those studied by geographers, are often driven by many interactions within and across scales that cannot be separated from one another (e.g., land cover; Vadjunec et al. 2018). Moreover, the processes responsible for generating the spatial patterns we seek to explain, and how those processes interact across scales, are likely to vary in time and space. For example, in a recent study investigating optimal scales for sampling atmospheric variables with unmanned aircraft systems, researchers found that the characteristic scales of temperature and humidity changed throughout the day and under different weather conditions (Hemingway et al. 2017). This combination of complexity and natural variation complicates R&R. Uncertainty about how to conceptualize the scale at which processes operate in replication studies conducted in different spatial locations is a fundamental challenge to R&R in geographic research. Even when location is held constant between studies, changes at scales not measured in an original study, and therefore omitted in a replication, can create failures to replicate that are not the result of the variables included in either study.

Still, decisions must be made regarding the scale at which to undertake geographical analyses, which leads to uncertainties in conceptualization (as well as measurement, discussed below). The most recognizable of these is the modifiable areal unit problem (MAUP; Openshaw 1984; Fotheringham and Wong 1991), which has plagued geographical analysis for decades. The early recommendation by Openshaw (1984) to collect data at the finest resolution possible is perhaps becoming more of a reality given improvements in technology for data capture. But, at the same time, individual data can lead to additional issues such as confidentiality. For example, in a recent project monitoring health impacts of transportation interventions in four cities, researchers grappled with tradeoffs between using nationally available data to ensure replicability across cities versus using detailed data to take advantage of the specific datasets collected by cities but at the expense of comparability between cities (Winters et al. 2018).

Measurement uncertainty

How we conceive geographic phenomena directly shapes how we measure those phenomena. However, when we measure reality there is also some uncertainty about how accurately and precisely our measurements match that reality. This measurement uncertainty can accompany both the locations and attributes of entities, and it typically arises midway through the scientific process during study design and data collection (Fig. 1). As with conceptual uncertainty, measurement uncertainty does not necessarily impact reproducibility, but it will affect replication if imperfect or unknown measurement conditions are not communicated. Errors in measurement and the uncertainties that surround them can be particularly problematic when they are spatially structured.

In geographical analysis, measurement uncertainties are often difficult to assess because geographers use *representations* of the real world that are almost always incomplete (Longley et al. 2015). Direct measurements of reality are often converted into representations using spatial data models that impose 'filters' on reality (Longley et al. 2015) and remove certain details. When this is the case, collecting and representing new data for replication studies following methods that match prior work can become difficult. Even basic decisions such as whether to represent a phenomena using a raster or vector data model introduce different tradeoffs between accuracy and precision. For example, Yoo and Trgovac (2011) found that changing the support of an analysis from point to interpolated raster for tree predictions led to discrepancies in the results. Spatial data uncertainty is also impacted by completeness and consistency (Haining 2009). One issue changing the interaction between measurement uncertainty and R&R is the emergence of new approaches for collecting spatial data. For instance, LiDAR technology allows the capture of 3D datasets with unprecedented accuracy and precision (e.g., Matasci et al. 2018). However, collection is not systematic across study areas or time periods, leading to issues of completeness and consistency that limit the potential for replication (Mathews et al. 2019).

The introduction of new forms of spatial data into geographic analysis has made it increasingly difficult to assess and communicate measurement uncertainties. For example, the rising prominence of citizen scientists and crowdsourcing within the geographic research process (Ferster et al. 2018) has introduced uncertainties related to participant reliability, instrumentation, calibration, and sampling that are difficult to quantify. Specifically, the normally opportunistic data collection methods of citizen science and related volunteered geographic information raise questions about the use of conventional forms of statistical inference due to the lack of representative sampling across space or time. Without a clear picture of what population is being studied, it is difficult to select appropriate distributions for comparison or develop informed priors. Without conventional statistical measures of data uncertainty or platforms for inference, it is difficult to make inferences about underlying processes, and by extension to reproduce or replicate those inferences. As one example, the BikeMaps.org project gathers volunteer-generated data for bicycle crashes, falls, and near misses that involve motor vehicles, scooters, pedestrians, infrastructure, or other cyclists (Nelson et al. 2015). Researchers have found that the official data typically capture about 30 percent of bicycle incidents in urban areas but are biased toward crashes that involve motor vehicles (Teschke et al. 2014). Without a complete map of bicycling safety data, it is difficult to quantify uncertainty or determine how VGI-based data collection may impact replicability.

As spatial data continue to expand in geographic coverage, collection frequency, and level of detail, it will be important for *Geographical Analysis* to lead the way in the development of standards and protocols for reporting the provenance of spatial data. Data provenance refers to the complete record of who did what to a piece of data, and how and why that adjustment was made. Standards and protocols of spatial data provenance are important because R&R require more than accessibility to an original spatial dataset; they require documentation of the modifications and transformations the data have undergone. To consider how researcher decisions about the construction of spatial data might be reasonably altered to fit a change in spatial context, or how such a change might affect replication, details and original reasoning about those choices should be made known. Tullis et al. (2015) and Linck (2015) suggest progress can be made in this area by developing methods to separate and quantify what they term 'content trust' related to the uncertainty in the data and data sources themselves, and 'workflow trust' related to the reliability of tools and algorithms used in data analysis. Tracking and communicating spatial data provenance becomes increasingly difficult in the collaborative, multi-user environments that increasingly typify geographic research because changes to data created by different users are often and easily lost during complex workflows.

Geographical Analysis can support the development of R&R by encouraging research that advances the development of provenance tracking frameworks (e.g., Singleton et al. 2016). Forum and standards for sharing data, code, and provenance information would similarly improve the R&R of geographic research. As a discipline, geography is well-positioned to test the effectiveness and limits of provenance frameworks for tracking and assessing measurement uncertainties because geographers have access to data sets measured consistently over long periods of time that also span a wide range of spaces. For example, the Landsat missions contain repeat data covering most of the globe since 1972 (Roy et al. 2014). Free, public access to Landsat data has led to broad usage, and while the 30m resolution is not ideal for measuring all phenomena, there are clear advantages to having a standard that can be used as a comparison for other measurements. Counter to the openness and availability of geographic data is the need to preserve privacy and limit data release to spatially aggregations (e.g., health records and movement tracking data). Thus, R&R is limited in certain contexts by the ability to record and share information about data provenance and measurement uncertainty.

Analytical Uncertainties

Uncertainty in the analysis of geographic data arises as researchers make choices about how to best examine spatial data given the natural variations, spatial patterns, and conceptual and measurement uncertainties that exist in the data. A goal of geographical analysis is to separate the noise present in spatial data from signals of the effects of spatial processes. The replication of spatial analysis is complicated by the number of choices that must be made during spatial analysis and the wide range of alternatives that could be reasonably pursued. As with conceptual and measurement uncertainties, tracking and communicating analytical uncertainties and the choices made to resolve them is fundamental to R&R in geography.

While it may be reasonable to expect a geographical analysis to be reproducible, it is not clear that all geographical analyses should be replicable. One of the fundamental features of geographical analysis is that it is not unusual to achieve different results about the determinants of a response variable based on data from different locations. This situation has two variants. First, different combinations of explanatory variables may appear to significantly affect the response in different locations. Second, the same model (i.e., combination of variables with same functional form) may be calibrated in different locations, but inferences regarding the variables affecting the response can vary. Local models are one form of geographical analysis that seek to assess variations in underlying spatial generating processes over space. Examples of such models include geographically weighted regression (Fotheringham et al. 1998, Fotheringham et al. 2002, Fotheringham et al. 2017, LeSage 2004), eigenvector spatial filter-based local regression (Griffith 2008, Oshan and Fotheringham 2017), and certain kinds of Bayesian spatially-varying coefficient models (Gelfand et al. 2003, Banerjee et al. 2014, Wolf et al. 2017). Using local models to analyze where models change in form or parameterization is akin to examining the locations in which a replication study may be expected to produce similar results and inferences.

If the goal is to produce regional expectations about the replicability of results, it is also necessary to consider the analytical uncertainties created when multiple tests are conducted in the presence of spatial data dependence. When multiple tests are used to examine the same hypothesis across a number of locations, it quickly becomes a near-certainty that results will include several false positives. Correction factors have been proposed (Bonferonni 1936, Sidak 1967, Benjamini and Hochberg 1995), but those corrections assume independence between tests that does not exist in most geographic contexts because neighboring test locations share spatially dependent data (Benjamini and Yekutieli 2001, Rogerson and Kedron 2012). In order to assess whether a replication that uses multiple testing produces similar results and inferences, it is necessary to first assess the effect of spatial dependence in the original study and the degree to which the data (and testing) used in the replication has similarly-structured spatial dependencies (de Silva and Fortheringham 2016, Rogerson and Kedron 2016). If the characteristics of the original data and methods are not communicated, replications may simply re-create and reinforce improper and incorrect inferences. Even when adjustments are made, analytical uncertainties remain because it is difficult to measure how well spatial dependencies were captured and incorporated into the corrections. This lack of knowledge further complicates assessment of the original inference and assessment of the R&R result.

Geographical Analysis can contribute to the advancement of R&R by supporting the development and use of practices that communicate researcher choice. However, improving communication through data/code sharing data and providing complete outputs, not just those presented in a final analysis, will not alone ensure the replicability of geographic research.

Because geographical analyses are directly tied to spatial data that are usually spatially autocorrelated it is often not possible to decompose datasets to isolate and examine selected characteristics. Small changes in location or geographic context can lead to important shifts in the spatial structure of data. When this occurs, we would not necessarily expect replication across space, but we can make progress on at least two fronts. First, we can use local forms of geographical analysis to search for consistencies in results across regions. Second, we can examine how 'context' is defined as well as what 'same' or 'similar' results mean in geography. Standardizing how these central ideas are defined as they relate to R&R in geography is key.

Of course, the issue of what is 'context' remains largely unanswered in geographical analysis (Robertson et al. 2018; Kwon 2012).

Communication Uncertainties

The communication of geographic research remains largely unidirectional and incomplete, which is perhaps the biggest challenge for R&R in geography. Geographic research is often presented as a linear process of conception, measurement, analysis, and results, which is then published as a peer-reviewed article, conference proceedings, research reports, or other form. However, such a clean pathway almost never occurs in reality, and the reality of what did occur is rarely fully reported in the paper. This practice creates problems not only for R&R but for research more broadly. Current publication culture dissuades inclusion of negative results due to either space and time constraints or a general lack of incentive (Rosenthal 1979, Franco et al. 2014). Therefore, negative results, which are equally important for advancing the state of geographic knowledge, are instead placed in the 'file drawer' where they are not disseminated to the larger scientific community. Similarly, while journals have begun to encourage authors to publish datasets and procedures accompanying their analysis, there are few incentives for researchers to adopt this practice (Munafo et al. 2017). Given that the development and preparation of a dataset can be time consuming, there is a strong disincentive for authors to provide open access before maximizing their own publications. These and other communication uncertainties exacerbate R&R by preventing and discouraging subsequent studies.

To facilitate the establishment of R&R in geography, *Geographical Analysis* can encourage researchers to provide access to their data/code and give clear descriptions of how the data were compiled and constructed. Singleton et al. (2016) go a step further in arguing that transparent workflows utilizing open source software and openly available data are necessary for ensuring reproducibility and scientific validity. While sharing data and code will help clarify how researchers manage measurement and analytical uncertainties, researchers should also explain choices made when conceptualizing and measuring the geographic phenomena under study.

Specific measures designed to address these conditions (e.g., registered reports, open science frameworks, calls for replication studies) are already being implemented in other scientific fields. A discussion of such tools applied to geographic research can be found in Nust et al. (2018). Incorporating those measures into the operation of the journal *Geographical Analysis* and, by extension, spatial analytical research is one avenue toward encouraging an R&R culture in our discipline.

Establishing R&R in *Geographical Analysis*

R&R has become a critical issue in many scientific fields, but geography and geographical analysis do not have a long history of addressing R&R. Some might argue that because geography is mostly a descriptive or methodological discipline and because geographers are not formally testing scientific hypotheses or making social policy recommendations, that R&R isn't particularly relevant for the discipline. We disagree. Improving explanation in geography is a primary goal of geographical analysis, and the underlying process of explanation is based on the scientific method, which is founded on the reproduction and reproducibility of results across different contexts. Regardless of whether a geographical analysis draws primarily from computational, theoretical, or experimental approaches, geographers should be working toward the reproduction and replication of *inference* across spatial contexts, because inferences that are robust to changes in data or spatial context will lead to the development of internally consistent and generalizable theory.

The complexity of spatial data makes our field vulnerable to criticism. Research examining the current degree to which existing work can be reproduced, and the space over which central findings are replicable, would position *Geographical Analysis* to respond to potential criticism and raise trust in the outcomes of geographic research. At the same time, further examination of the practice of geographical analysis itself as it relates to uncertainties, the characteristics of spatial data, and the sharing of results will help resolve how replicable we can expect geographic research to be across space. In this way, research into R&R can reinvigorate a foundational debate within geography between the nomothetic and idiographic approaches of the discipline. As a practical step forward, expanding research into geographic ontologies, frameworks for spatial data provenance, and the efficient sharing of all analytical choices and results can improve our present ability to reproduce and replicate geographical analyses and to study R&R. Sharing data and code are necessary, but alone these steps are not sufficient to ensure R&R in geographical analysis. The structured, but not entirely predictable, spatial variation of geographic phenomena means that replicability in geographical analysis may not be an obtainable objective.

Nonetheless, we argue that developing a culture of R&R in geography is the best way to facilitate the rigorous examination of why regularities in spatial patterns and processes breakdown and how spatial context shapes those changes.

References

Baker, M. (2015). "Over half of psychology studies fail reproducibility test." *Nature News*. Available at https://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248 Accessed October 30, 2018.

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

Banerjee, S., B. P. Carlin, and Gefland, A.E. (2014). *Hierarchical modeling and analysis for spatial data*. London: CRC Press.

Barba, L. A. 2018. "Terminologies for reproducible research." Available at http://arxiv.org/abs/1802.03311 Accessed February 9, 2018.

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society B* 57(1), 289-300.

Benjamini, Y., and Yekutieli, D. (2001). "The control of the false discovery rate in multiple testing under dependency." *The Annals of Statistics*, 29(4), 1165-1188.

Bollen, K., Cacioppo, J.T., Kaplan, R. Krosnick,J. and Olds, J.L. (2015). "Social, Behavior, and Economic Sciences Perspectives on Robust and reliable Science (National Science Foundation, Arlington, VA).

Bonferroni, C.E. (1936). "Teoria statistica delle classi e calcolo delle probabilita." *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze* 8, 3-62.

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

Brunsdon, C. and Singleton, A. (2015). "Reproducible research: Concepts, techniques and issues." In *Geocomputation: A Practical Primer* 254-264, edited by C. Brunsdon and A. Singleton. London: Sage.

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

Claerbout, J. F., and Karrenbach, M. (1992). “Electronic documents give reproducible research a new meaning.” *SEG Expanded Abstracts* 11, 601–604.

da Silva, A. and Fotheringham, A. S. (2016). “The Multiple Testing Issue in Geographically Weighted Regression.” *Geographical Analysis* 48(3), 233-247.

de Groot, A. D. (2014). The meaning of “significance” for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han LJ van der Maas]. *Acta psychologica* 148, 188-194.

Donoho, D. L., Maleki, A., Rahman, I.U., Shahram, M., and Stodden, V. (2009). “Reproducible research in computational harmonic analysis.” *Computing in Science & Engineering*, 11(1), 8–18

Ferster, C.J., Nelson, T.A., Robertson, C. and Feick, R. (2018). “Current Themes in Volunteered Geographic Information.” in *Comprehensive Geographic Information Systems*. 26-41, edited by B. Huang. Elsevier.

Fotheringham, A.S., C. Brunsdon, and Charlton, C. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, London:Wiley.

Fotheringham, A. S., M. E. Charlton, and Brunsdon, C. (1998). “Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis”. *Environment and Planning A* 30(11): 1905–1927

Fotheringham, A.S., W. Yang and Kang, W. (2017) “Multiscale geographically weighted regression (MGWR).” *Annals of the American Association of Geographers* 107(6): 1247-1265.

Fotheringham, A.S. and Wong, D. W-S. (1991). “The modifiable areal unit problem in multivariate statistical analysis.” *Environment and Planning A* 23(7), 1025-1044.

Franco, A., Malhotra, N. and Simonovits, G. (2014). "Publication bias in the social sciences: Unlocking the file drawer." *Science* 345 1502–1505.

Gelfand, A. E., H-J Kim, C.F. Sirmans, and Banerjee, S. (2003). "Spatial modeling with spatially varying coefficient processes." *Journal of the American Statistical Association* 98 (462), 387–396.

Gelman, A. and Loken, E. (2014). "Ethics and statistics: The AAA tranche of subprime science." *Chance* 27(1), 51-56.

Goodchild, M.F. (2018). "A GIScience Perspective on the Uncertainty of Context." *Annals of the American Association of Geographers* 108(6), 1476-1481.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). "What does research reproducibility mean?." *Science translational medicine*, 8(341), 341ps12-341ps12.

Griffith, D.A. (2008). "Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR)." *Environment and Planning A* 40(11), 2751–2769

Griffith, D.A. (2018). "Uncertainty and Context in Geography and GIScience: Reflections on Spatial Autocorrelation, Spatial Sampling, and Health Data." *Annals of the American Association of Geographers* 108(6), 1499-1505.

Haining, R. (2009). "The special nature of spatial data." In *The Sage Handbook of Spatial Analysis* 5-24, edited by A.S. Fotheringham and P.A. Rogerson. London: Sage

Hartshorne, R. (1939). "The nature of geography: A critical survey of current thought in the light of the past." *Annals of the Association of American Geographers* 29(3), 173-412.

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

Hemingway, B.L., A.E. Frazier, B.R. Elbing, and Jacob, J.D. (2017). "Vertical sampling scales for atmospheric boundary layer measurements from small unmanned aircraft systems (sUAS)." *Atmosphere*, *8*(9), 176.

Ioannidis, J.P. (2005). “Why most published research findings are false.” *PLoS medicine*, 2(8), e124.

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

Konkol, M., Kray, C., and Pfeiffer, M. (2019). “Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study.” *International Journal of Geographical Information Science*, 33(2), 408-429.

Kwan, M.P. (2012). “The uncertain geographic context problem.” *Annals of the Association of American Geographers 102*(5), 958-968.

Linck, R.A. (2015) "Geospatial Workflows and Trust: a Use Case for Provenance.” Theses and Dissertations. 1364. https://scholarworks.uark.edu/etd/1364

LeSage, J. P. (2004). “A Family of Geographically Weighted Regression Models.” in *Advances in Spatial Econometrics*, 241-264, edited by L. Anselin, R. Florax, and S.J. Rey. Berlin:Springer.

Longley, P.A., M.F. Goodchild, M.F, D.J. Maguire, and Rhind, D.W. (2015). *Geographic Information Science and Systems*. Hoboken, NJ, USA: John Wiley & Sons.

Matasci, G., T. Hermosilla, M.A. Wulder, J.C. White, N.C. Coops, G.W. Hobart, and Zald, H.S. (2018). “Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots.” *Remote Sensing of Environment*, 209, 90-106.

Mathews, A.J., A.E. Frazier, S.V. Nghiem, G. Neumann, and Zhao, Y. (2019). Satellite scatterometer estimation of urban built-up volume: validation with airborne lidar data. International Journal of Applied Earth Observation and Geoinformation, 77, 100-107.

Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Du Sert, N.P., Simonsohn, U., Wagenmakers, E.J., Ware, J.J. and Ioannidis, J.P. (2017). “A manifesto for reproducible science”. *Nature Human Behaviour*, 1(1), 0021.

Nelson, T.A., Denouden, T., Jestico, B., Laberee, K. and Winters, M. (2015). "BikeMaps.org: A global tool for collision and near miss mapping." *Frontiers in Public Health* 3, 53.

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

Openshaw, S. (1984). "The modifiable areal unit problem." *Concepts and Techniques in Modern Geography*. 38. Norwich, U.K., GeoBooks.

Oshan, T. and Fotheringham, A.S. (2017). "A comparison of spatially varying coefficient estimates using geographically weighted and spatial-filter-based techniques." *Geographical Analysis* 50(1), 53-75.

Ostermann, F.O. and Granell, C. (2017). "Advancing science with VGI: Reproducibility and replicability of recent studies using VGI." *Transactions in GIS*, 21(2), 224-237.

Peng, R. (2015). "The reproducibility crisis in science: A statistical counterattack." *Significance* 12(3), 30-32.

Plesser, H. E. (2018). "Reproducibility vs. replicability: a brief history of a confused terminology." Frontiers in neuroinformatics, 11, 76.

Robertson, C., and Feick, R. (2018). "Inference and analysis across spatial supports in the big data era: Uncertain point observations and geographic contexts." *Transactions in GIS 22*(2), 455-476.

Rogerson, P.A. and Kedron, P.J. (2012). "Optimal weights for focused tests of clustering using the Local Moran statistic." *Geographical Analysis* 44(2), 121-133.

Rogerson, P.A. and Kedron, P.J. (2016). "A changepoint statistic with uniform type I error probabilities." *Communications in Statistics-Theory and Methods* 45(16), 4663-4672.

Rosenthal, R. (1979). "File drawer problem and tolerance for null results." *Psychological Bulletin* 86(3), 638–641.

Roy, D.P., M.A. Wulder, T.R. Loveland, C.E. Woodcock, R.G. Allen, M.C. Anderson, D. Helder, J.R. Irons, D.M. Johnson, R. Kennedy, and Scambos, T.A. (2014). “Landsat-8: Science and product vision for terrestrial global change research.” *Remote Sensing of Environment* 145(5), 154-172.

Schaefer, F. K. (1953). “Exceptionalism in geography: A methodological examination.” *Annals of the Association of American Geographers* 43(3), 226-249.

Šidák, Z. K. (1967). "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions". Journal of the American Statistical Association. 62 (318): 626–633.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science* 22(11), 1359-1366.

Singleton, A.D., Spielman, S. and Brunsdon, C. (2016) Establishing a framework for Open Geographic Information Science. *International Journal of Geographical Information Science,* 30(8), 1507-1521.

Smith, B. and Mark, D.M. (2001). “Geographical categories: an ontological investigation.” *International Journal of Geographical Information Science*. 15(7), 591-612.

Smith, B. and Mark, D.M. (2003). “Do mountains exist? Towards an ontology of landforms.” *Environment and Planning B: Planning and Design*, 30(3), 411-427.

Stodden, V. (2015). “Reproducing statistical results.” *Annual Review of Statistics and Its Application* 2, 1-19.

Teschke, K., T. Frendo, H. Shen, M.A. Harris, C.C. Reynolds, P.A. Cripton, J. Brubacher, M.D. Cusimano, S.M. Friedman, G. Hunte, and Monro, M. (2014). “Bicycling crash circumstances vary by route type: a cross-sectional analysis.” *BMC Public Health* 14(1), 1205.

Tullis, J., Cothren, J., Lanter, D., Shi, X., Limp, F., Linck, R., Young, S., Alsumaiti, T. Geoprocessing, Workflows, and Provenance. *Remote Sensing Handbook*. 1st ed. Vol. 1. Taylor and Francis.

Vadjunec, J.M., Frazier, A.E., Kedron, P., Fagin, T. and Zhao, Y., 2018. A Land Systems Science Framework for Bridging Land System Architecture and Landscape Ecology: A Case Study from the Southern High Plains. *Land*, *7*(1), p.27.

Winters, M., M. Branion-Calles, S. Therrien, D. Fuller, L. Gauvin, D.G. Whitehurst, and Nelson, T.A. (2018). "Impacts of Bicycle Infrastructure in Mid-Sized Cities (IBIMS): protocol for a natural experiment study in three Canadian cities." *BMJ Open* 8(1), p.e019130.

Wolf, L. J., T.M. Oshan, and Fotheringham, A.S. (2017). "Single and multiscale models of process spatial heterogeneity." *Geographical Analysis* 50(3), 223-246.

Yoo, E.H. and Trgovac, A.B. (2011). "Scale effects in uncertainty modeling of presettlement vegetation." *International Journal of Geographic Information Science* 25(3), 405-421.

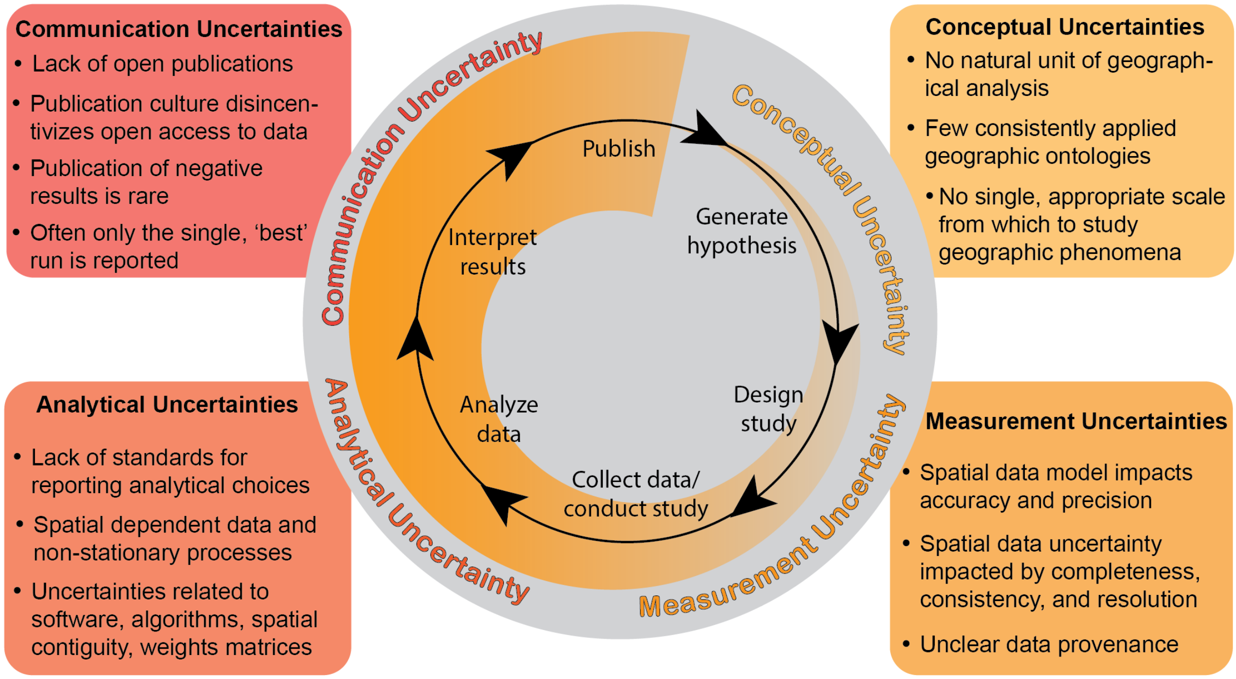

Figure 1. Four geographic uncertainty filters coupled with the scientific method. The boxes provide examples of uncertainty sources that complicate the replication and reproduction of geographic research. Adapted from Munafo et al. (2017).