A Vision for Verdical Spatial Data Science *

Peter Kedron[1,2][0000−0002−1093−3416], Sarah Bardin[1,2][0000−0001−8657−1725]

[1] Arizona State University, Tempe AZ 85281, USA
[2] Spatial Analysis Research Center, Tempe AZ 85283, USA
Peter.Kedron@asu.edu

Abstract. To hold the same privileged epistemological position as science, spatial data science must satisfy the self-corrective thesis. Doing so depends on the field's capacity to reproduce and replicate published work, the willingness of researchers to do so, and our ability to assess the cumulative insights of such studies. We present some steps spatial data science might take to develop these capabilities and put forward a provisional vision of a veridical spatial data science.

Keywords: Reproducibility · Replication · Self-correction in Science · Evidence Accumulation · Spatial Data Science.

DOI: https://doi.org/10.25436/E20016

1 Self-correction in Science and Spatial Data Science

To hold the same privileged epistemological position as science, Spatial Data Science (SDS) must satisfy Peirce's (25) self-corrective thesis (SCT)[1].

Theorem 1. *In the long run, the scientific method will refute false theories and find closer approximations to true theories.*

The SCT is theoretically plausible in conventional science because the scientific method allows for error correction through the reproduction, replication, and the repeated testing of beliefs about empirical phenomena[2]. Crucially, the scientific method recognizes that the data we use to make claims about the world are unstable and confounded by aleatoric uncertainty and the potential influence of contextual factors. To mitigate the influence those confounds have on our understanding of the world, scientists look for recurrent patterns across multiple data sets gathered in different contexts to build theories about the phenomena they believe to be the causes of those patterns (6; 32).

Establishing belief in scientific theories relies on reproductions and replications to check the internal and external validity of claims about phenomena, respectively (18; 9). Reproductions assess the internal validity of research claims by repeating the procedures of an original analysis using the same data, or by modifying those procedures to test the robustness of the original claims to

* This material is based on work supported by the National Science Foundation under Grant No. Spatial (**BCS-2049837**)

perturbation. In this way, a reproduction study checks the credibility of the procedures used to (a) link observed data to unobserved phenomena and (b) make claims about those phenomena. Forming those claims into a theoretical explanation of phenomena rests on the external validity check offered by replication. A replication study uses new data to retest the claims made in a prior analysis. Repeated across time and space, replications allow us to separate phenomena from the variability of data and eventually identify and and better understand the influence of context and to infer the stability of phenomena. Those stable, recurrent features of the world are what we hope to explain through theory.

Whether the SCT is true in science or SDS depends on (a) researcher practices, (b) whether researchers are attempting reproductions and replications of existing work, and (c) how well approaches to knowledge accumulation are functioning. At present, many spatial data scientists continue to publish irreproducible work (24; 11), formal reproductions and replications of spatial analyses remain rare (1; 8), and limited research effort has gone into developing formal systems of knowledge accumulation that recognize the unique challenges that accompany spatial data analysis (5; 8). Advancing SDS as a science requires addressing these challenges.

2 Placing the Cornerstones of Spatial Data Science

Recent research in geography and data science has laid some of the conceptual (8; 27; 33), technical (29; 22; 31), and practical foundations (9; 3; 12) needed to advance SDS as a science. Yu and Kumbier's (33) call for the development of a veridical data science (VDS) neatly aggregates key aspects of this literature. Specifically, the authors define VDS as,

> "**Veridical Data Science (VDS)**principled inquiry to extract reliable and reproducible information from data, with an enriched technical language to communicate and evaluate empirical evidence in the context of human decisions and domain knowledge."

VDS not only incorporates but goes beyond typical calls to share data and code or improve research provenance through forms of literate programming by also incorporating an internal mechanism to check the stability of results to specification decisions made throughout a data science project. Like other forms of multiverse analysis, the perturbation analysis proposed by Yu and Kumbier is intended to explore the complete set of plausible ways a study could have been executed, and as a result checks of the internal validity of a finding. The shortcoming of the VDS approach, and similar frameworks (4), is that it does not directly incorporate a parallel plan to systematically check external validity.

While replication studies evaluate the external validity of a study, individual replications need to be considered together to reduce the influence of confounds on the measurement and understanding of phenomena. Meta-analyses (7; 14) that subject collections of published results to secondary analysis are one means of assessing the cumulative evidence provided by both novel and replication studies. In a typical meta-analysis the results of many studies are formally aggregated to make an overall inference about the external validity of a claim.

However, such meta-analyses are often difficult to execute in practice and can suffer from selection and interpretation biases (30). Moreover, meta-analyses are retrospective in orientation and opportunistic in nature, which limits the ability to design prospective sequences of replications that can act as external validity checks across contexts.

To address these issues, Nichols et al. (19; 20) advocate an analogous information state approach (ISA) to evidence accumulation in which the predictive performance of m competing models across a sequence of studies is used to iteratively adjust a vector of weights, $\pi_t(model_i)$, that reflects the level of confidence we have in each model at time t. As data from a new replication conducted at $t+1$ become available, their consistency with each model can be assessed through the application of Bayes Theorem - combining the prior weight vector with the new data to create a new weights vector representative of the present information state. Nichols approach is analogous to that of Burnham and Anderson (4) who argue that information criterion, such as AIC, can be used to obtain savvy model weights to support multimodel inference as opposed to choosing the single "best" model specification from a group of alternatives.

The critical benefit of these approaches is that they theoretically present the opportunity to create prospective sequences of replications that can be designed to discriminate among competing models, and their associated claims/hypotheses. Adopting these changes in practice, and perspective, would allow us to integrate external validity checks directly into the design of our research programs, thereby increasing the chances of our research fulfilling the SCT.

3 A Vision of a Verdical Spatial Data Science

Building on the cornerstones of VDS, ISA, and our existing knowledge of spatial analysis, we believe that it is possible to establish a veridical spatial data science (VSDS) capable of satisfying the SCT.

We can build a reproduction-based check of the *internal validity* of individual spatial analyses by adapting VDS to the study of spatial data and geographic phenomena. Converting VDS will be an extended process that will require us to build provenance systems designed for spatial data and incentive systems that encourage reproduction. Within the geographical sciences, the work to create these systems is already underway. The NSF-funded CyberGIS Center (29) and the Open Geospatial Consortium (OAC-1743184) have developed cyberinfrastructure and geospatial software standards to pursue computationally reproducible research, while the Opening Reproducible Research project is developing publication standards and creating software to facilitate data, code, and computational environments (21; 22; 23). At the same time, recent publications have linked reproduction and replication to the geography's core traditions (28), methodological approaches (1; 2; 8), educational priorities (17), and enduring theoretical debates (27).

Another productive step would be to modify perturbation analysis to track how responsive effect estimates are to decisions that are explicitly spatial in nature (e.g., definitions of spatial extent, support, relationships) and then use this information to assign a portion of the range of confidence we have in an effect estimate to those spatial decisions. Kedron et al's. work linking reproducibility issues to the spatial analysis lifecycle (8) and the use of specification curves in spatial data analysis (10) has laid the groundwork for this type of modification.

We can similarly adapt Nichols et al's. (19) information state approach to the study of spatial data to design prospective sequences of replications that both check the *external validity* of our claims and provide the evidence we need to adjust our beliefs in competing hypotheses. Goodchild and Li (5) argue that replication across space will always be weak, but that does not mean that replications cannot provide information useful in the differentiation of competing hypotheses. A central challenge to designing study sequences and creating statistical approaches to updating the information state will be addressing how geographic context, spatial non-stationarity, and spatial dependence confound the information we receive from replications undertaken in different locations and at different geographic scales.

One approach to this challenge is to design series of integrated experiments that use the variety of evidence[3] place-based differences provide to first recognize an effect of interest and then identify the place and conditions in which that effect can be reliably observed. Further adapted to space and place, List's (15) proposed approach to testing the scalability of policy interventions offers an intriguing path forward. Specifically, List calls for the development of a research hierarchy that involves three waves of studies. In a first wave, a proof of concept analysis is conducted to establish belief in a hypothesis under ideal conditions. A second wave of studies tests the hypothesis in varying contexts to establish boundary conditions. A final wave of studies then uses multi-site experiments to measure change in the magnitude of effects across contexts. Modifying this approach to account for geographic scale and spatial dependence in the second and third waves is one potential avenue towards establishing the domain of theories that hold across some, but not all, geographic contexts and scales.

Two related practical challenges will be (a) convincing funding agencies to shift their focus to support the accumulation of knowledge and the integration of results and (b) training and incentivizing researchers to design and execute replication sequences that will require increased collaboration, resources, and time to complete. An effort capable of addressing both of these issues simultaneously would be the establishment of an NSF Research Traineeship Program in VSDS. By training a generation of scholars in replication and working with those researchers to build the basic science of knowledge accumulation across space, we might well create the culture of collaboration needed to support VSDS.

We believe the steps above can begin the transition to VSDS, which we provisionally define as:

> **Veridical Spatial Data Science (VSDS)** - a principled inquiry to extract reliable and reproducible information from *spatial data*, with an enriched technical language to communicate and evaluate empirical evidence in the context of human decisions, domain knowledge, and *geographic confounds; and a system of external validation and evidence accumulation based on the purposeful replication of findings across space.*

Notes

[1]Here we state Romero's (26) formulation of the SCT. More formally the SCT concerns the justification of inductive inference in science. Mayo (16) states the SCT as, *"Methods for inductive inference in science are error correcting; the justification for inductive methods of experimental testing in science is that they are self-correcting.* We focus on the key implication of the thesis - that no matter where different researchers begin their investigation, if they follow the scientific method, their results will eventually converge toward the same outcome, the truth.

[2]Whether science itself meets this same thesis remains an open philosophical debate and is an issue at the center of the present reproducibility crisis

[3]More broadly, the variety of evidence thesis claims that, ceteris paribus, varied evidence has higher confirmatory power than less varied evidence (13)

Bibliography

[1] Brunsdon, C.: Quantitative methods i: Reproducible research and quantitative geography. Progress in Human Geography **40**(5), 687–696 (2016)

[2] Brunsdon, C., Comber, A.: Opening practice: supporting reproducibility and critical spatial data science. Journal of Geographical Systems pp. 1–20 (2020)

[3] Brunsdon, C., Singleton, A.: Geocomputation: a practical primer. Sage (2015)

[4] Burnham, K.P., Anderson, D.R.: Multimodel inference: Understanding aic and bic in model selection. Sociological methods research **33**(2), 261–304 (2004)

[5] Goodchild, M.F., Li, W.: Replication across space and time must be weak in the social and environmental sciences. Proceedings of the National Academy of Sciences **118**(35), e2015759118 (Aug 2021). https://doi.org/10.1073/pnas.2015759118, http://www.pnas.org/lookup/doi/10.1073/pnas.2015759118

[6] Haig, B.: Understanding replication in a way that is true to science (2020)

[7] Hedges, L.V., Olkin, I.: Statistical methods for meta-analysis. Academic press (2014)

[8] Kedron, P., Frazier, A.E., Trgovac, A.B., Nelson, T., Fotheringham, A.S.: Reproducibility and replicability in geographical analysis. Geographical Analysis **53**(1), 135–147 (2021)

[9] Kedron, P., Li, W., Fotheringham, S., Goodchild, M.: Reproducibility and replicability: opportunities and challenges for geospatial research. International Journal of Geographical Information Science **35**(3), 427–445 (2021)

[10] Kedron, P., Quick, M., Hilgendorf, Z., Sachdeva, M.: Using the specification curve to teach spatial data analysis and explore geographic uncertainties. Journal of Geography in Higher Education pp. 1–11 (2021)

[11] Konkol, M., Kray, C., Pfeiffer, M.: 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 (2019)

[12] Konkol, M., Nüst, D., Goulier, L.: Publishing computational research-a review of infrastructures for reproducible and transparent scholarly communication. Research integrity and peer review **5**(1), 1–8 (2020)

[13] Landes, J.: Variety of Evidence. Erkenntnis **85**(1), 183–223 (Feb 2020). https://doi.org/10.1007/s10670-018-0024-6, https://doi.org/10.1007/s10670-018-0024-6

[14] Lipsey, M.W., Wilson, D.B.: Practical meta-analysis. SAGE publications, Inc (2001)

[15] List, J.A.: Non est disputandum de generalizability? a glimpse into the external validity trial. Working Paper 27535, National Bureau of Economic

Research (July 2020). https://doi.org/10.3386/w27535, http://www.nber.org/papers/w27535

[16] Mayo, D.G.: Peircean induction and the error-correcting thesis. Transactions of the Charles S. Peirce Society **41**(2), 299–319 (2005)

[17] Muenchow, J., Schäfer, S., Krüger, E.: Reviewing qualitative gis research—toward a wider usage of open-source gis and reproducible research practices. Geography Compass **13**(6), e12441 (2019)

[18] NASEM: Reproducibility and Replicability in Science. National Academies Press (2019)

[19] Nichols, J.D., Kendall, W.L., Boomer, G.S.: Accumulating evidence in ecology: Once is not enough. Ecology and Evolution **9**(24), 13991–14004 (Dec 2019). https://doi.org/10.1002/ece3.5836, https://onlinelibrary.wiley.com/doi/10.1002/ece3.5836

[20] Nichols, J.D., Oli, M.K., Kendall, W.L., Boomer, G.S.: Opinion: A better approach for dealing with reproducibility and replicability in science. Proceedings of the National Academy of Sciences **118**(7) (Feb 2021). https://doi.org/10.1073/pnas.2100769118, https://www.pnas.org/content/118/7/e2100769118

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

[22] Nüst, D., Hinz, M.: containerit: Generating dockerfiles for reproducible research with r. Journal of Open Source Software **4**(40), 1603 (2019)

[23] Nüst, D., Ostermann, F., Hofer, B., Granell, C., Sileryte, R.: Reproducible research at agile (2021), https://reproducible-agile.github.io/

[24] Ostermann, F.O., Granell, C.: Advancing science with vgi: Reproducibility and replicability of recent studies using vgi. Transactions in GIS **21**(2), 224–237 (2017)

[25] Peirce, C.S.: On the logic of drawing history from ancient documents, especially from testimonies. The Essential Peirce, 1893-1913 **2**, 75–114 (1901)

[26] Romero, F.: Philosophy of science and the replicability crisis. Philosophy Compass **14**(11), e12633 (2019)

[27] Sui, D., Kedron, P.: Reproducibility and replicability in the context of the contested identities of geography. Annals of the American Association of Geographers **111**(5), 1275–1283 (2021)

[28] Wainwright, J.: Is critical human geography research replicable? Annals of the American Association of Geographers pp. 1–7 (2020)

[29] Wang, S.: Cybergis and spatial data science. GeoJournal **81**(6), 965–968 (2016)

[30] Williamson, P., Gamble, C.: Identification and impact of outcome selection bias in meta-analysis. Statistics in medicine **24**(10), 1547–1561 (2005)

[31] Wilson, J.P., Butler, K., Gao, S., Hu, Y., Li, W., Wright, D.J.: A five-star guide for achieving replicability and reproducibility when working with gis software and algorithms. Annals of the American Association of Geographers **111**(5), 1311–1317 (2021)

[32] Woodward, J.F.: Data and phenomena: a restatement and defense. Synthese **182**(1), 165–179 (2011)

[33] Yu, B., Kumbier, K.: Veridical data science. Proceedings of the National Academy of Sciences **117**(8), 3920–3929 (Feb 2020). https://doi.org/10.1073/pnas.1901326117, https://www.pnas.org/content/117/8/3920