Vision Statement

Spatial Data Infrastructure to Identify What Works Where

Peter Kedron

The Vision

> *GIScientists can contribute to climate change mitigation and adaptation by creating the spatial data analysis and conceptual infrastructure needed to evaluate theories of change and scale policy intervention across locations.*

Actions to mitigate or adapt to climate change are ultimately interventions enacted in specific spatial and temporal contexts to address the effects of shifts in the climate. Those interventions are based on theories that explain why the climate is shifting and predict how the human-environment system will react to a particular action. While the scientific community has high confidence in the causes of climate change, the effectiveness of adaptations and mitigations has only been documented in specific contexts, sectors, and regions (IPCC 2023). It is not yet clear which mitigation and adaptation strategies will work best in which social and environmental situations. This uncertainty creates a clear need to test interventions across locations and collectively evaluate the information derived from those tests. This need will only rise as climate change accelerates and produces geographically varied shifts in social and environmental conditions. Until the scientific community can provide reliable evidence identifying which climate interventions are likely to work in which local context, policymakers will have limited guidance on when and where to implement adaptation and mitigation strategies which could result in the misallocation of resources and the exacerbation of inequalities.

GIScientists can help build the evidence base needed to support predictions about the effectiveness of climate interventions across locations in at least two ways. First, GIScientists can build spatial data infrastructure for effectiveness predictions. This task requires the integration of the core of GIScience, conceptualizing and creating systems to collect, store, and analyze spatial data (Mark 2003), with the synthesis approach of a human-environment centered geography that examines phenomena within a web of complex processes interacting within and between locations (see Turner 1989, 2002). Crucially, working within a synthesis science allows geographers and GIScientists to not simply measure and catalog policy impacts across locations, but to identify the set of factors that need to be in place for an intervention to work as expected in another location. Following disciplinary traditions, geographers can take theories of climate adaptation and mitigation developed in other fields, root them in place(s), and analyze how localized variations in process relationships impact the sufficiency of those explanations. Leveraging the information produced during investigations of different locations, geographers can create empirically grounded, middle-range theories (Harvey 1969, Miller and Goodchild 2015) that can act as stepping stones to the development of more general explanations of the processes that shape climate change mitigation and adaptation. This process functions best if an infrastructure to gather and analyze spatial data related to climate interventions is in place and in use.

Second, while the development of middle-range theories can support climate action, a conventional approach to scientific discovery centered on expectations of regularity, controlled experimentation, and slow progressive consensus building may be ill-suited to inform implementation of short-term climate interventions. Instead, because climate change presents a high stakes and uncertain challenge in need of immediate action, it is essential to emphasize evidential quality rather than scientific certainty (Bray and von Storch 1999, Kraus et al. 2012). Building on Functowicz and Ravetz (1993), high quality evidence will provide information about the feasibility and viability of interventions across technical and social constraints that are likely to vary from location to location. However, assessing quality in the uncertain context of climate change will also require the cultivation and sustained engagement of an extend peer community capable of bringing forms of expertise and knowledge from outside conventional science to bear on effectiveness predictions. GIScientists can contribute to the development of this peer community by building on the field's critical and participatory traditions. Doing so should not only improve intervention evaluation, but bring forward local inputs about what aspects of the environment and society should be changed, and how and where those changes should be made.

To address these challenges, we propose GISphere-KG, an AI-powered platform utilizing the Knowledge Graph (KG) and advanced large language models (LLMs) through LangChain and Neo4j. KG, with its ability to organize diverse data including structured and unstructured data (Li et al., 2022), helps in efficient information retrieval (Li et al., 2023), similar research interest discovery, and answering queries relevant to GIS education. LLMs, capable of semantic recognition for entities and relationships, and intent detection, convert natural language into machine-understandable statements used by graph database (i.e. Neo4j) to support tasks mentioned above. This integration aims to offer a comprehensive, user-friendly tool that not only provides up-to-date GIS program information worldwide but also uses semantic similarity calculations to match applicants with professors and programs fitting their research interests and preferences. GISphere-KG's innovative approach has the potential to enhance global GIS education. Furthermore, such a useful platform can be extended to support search and recommendation for other kinds of expertise by mining diverse source of data.

Identifying What Works Where

Researchers in other disciplines specialize in producing evidence that a mitigation or adaptation strategy has or has not worked in a particular location at a particular time (Grace 2017). Usually employing some form of randomized control trial (RCT) or quasi-experimental design (QED), these scholars gather high quality evidence that establishes the efficacy of an intervention. However, evidence that an intervention had a desired effect in one location is, in and of itself, of limited use when trying to predict whether that same intervention will have a similar effect in another location. To make this prediction about the effectiveness of the intervention across locations, it is also necessary to gather additional facts that establish that a reliable, systematic connection exists between the policy and the effect, and facts that identify the support factors that must be in place for that connection to function as predicted (see Cartwright 2007, Cartwright and Hardie 2012). Which facts are relevant to the effectiveness prediction, and will ultimately count as evidence, depend on the causal mechanism and theoretical reasoning used to formulate the intervention. Once elaborated, gathering and verifying those facts becomes a matter of empirical practice.

Synthesis of the form geographers specialize in is essential to gathering the facts needed to support an effectiveness prediction. However, the current policy evaluation environment is designed to establish the efficacy of interventions in particular locations. Focused on efficacy, researchers often do not design studies to gather the additional information needed to predict effectiveness across locations. The results of current practices can be observed in the catalogs of evidence clearinghouses and meta-regressions of policy impacts. While evidence clearinghouses contain numerous high quality policy evaluations, they do not systematically catalog basic spatial attribute data such as the exact location and scale of an intervention, much less the multitude of other local attributes that might impact how a policy functions in a different context. Meta-regressions use the evidence generated across multiple impact evaluations to make pooled estimates of the average impact of an intervention. However, these analyses also typically assume that policy impacts are derived from spatially stationary processes and that evaluations conducted in proximate locations have no effect on one another. These assumptions are unlikely to be true, which can affect the reliability of the impact estimates. Moreover, because they are averaged across locations the impact estimates meta-regressions produce provide limited insight into how location- specific factors affect the efficacy of an intervention.

Evaluation practices focused on measuring the magnitude of intervention effects may also be poorly aligned with the urgency of the climate challenge. When a large amount of uncertainty exists, but there is nonetheless an urgent need to act, precise estimates of the magnitude of an intervention effect are less useful than a clear signal of the direction of the effect and a sound understanding of the factors that allow that effect to occur. Relatedly, there is a need to arrive at this understanding quickly across locations, which is likely to be done best by leveraging the knowledge of a wider set of contributors that includes those with local-, topical-, and system-related expertise. The challenge, as ever, is creating systems to gather and integrate those perspectives. Beyond the traditions of GIScience and geography, the evolving fields of decision ad implementation science offer insights and practices to build upon.

Creating Spatial Infrastructure to Identify What Works Where

GIScientists can begin to support the evaluation of climate mitigation and adaptation strategies and the transfer of successful interventions across locations by improving the basic spatial data infrastructure of evidence collection and analysis. GIScientists can educate and assist researchers conducting policy evaluations to collect location data about the spatial extent of an intervention and other attributes predicted to influence the causal effect of the intervention by its underlying theory. Ideally, this data would be accompanied by spatial metadata that conforms to standards set out by the Open Geospatial Consortium

(OGC 2023) and the Federal Geographic Data Committee (FGDC 2022) to facilitate reuse and reproducibility.

Building such as spatial data infrastructure is a non-trivial objective for several reasons. First, it is typically not clear what the geographies of an adaptation or mitigation intervention are. Climate interventions are often composed of multiple actions that occur at multiple, overlapping scales, which makes identifying even the spatial extent of an action challenging. Second, it is also unclear how to best capture some interventions as spatial data. For example, how should a system of infiltration ditches and the effect they produce on flooding be recorded as spatial data? Is it better to represent the individual ditches as line features, or to record the intervention as change in the larger drainage network?

Correspondingly, should the effect of the intervention be recorded immediately surrounding the new ditches, or as a shift in the function of the larger network? Third, even if these questions are addressed, it is also unclear if current standards can capture intervention data. A research space exists for GIScientists to modify existing data standards and evidence protocols to the task of effectiveness prediction.

Beyond adoption in individual evaluations, the GIScience community could lead an effort to develop and improve inventories of adaptation plans and evidence clearinghouses containing impact evaluations. A comprehensive catalog on climate interventions and impact evaluations has yet to be developed and may be an unattainable goal, but a clearinghouse with even partial coverage would help policymakers and researchers examine efficacy and effectiveness. Building such a clearinghouse would also require the development of an evidence hierarchy for effectiveness predictions. Evidence hierarchies of existing clearinghouses identify meta-analyses and RCTs as the highest form of evidence because they prioritize measurements of efficacy. However, effectiveness predictions are only partially informed by the outcomes of these designs. An effectiveness-focused hierarchy would also need to prioritize spatial information about the factors that support the function of an intervention, so researchers can work to find similar settings for future interventions. However, identifying these factors must be done in relation to the theory underlying the interventions and local knowledge of the systems they exist in, which makes development of a cross-cutting standard difficult.

Finally, collectively analyzing impact evaluations of climate policies across regions and identifying criteria useful for effectiveness predictions is, by definition, a spatial analytical task. Presently, meta- regressions of climate policy impacts rest on questionable assumptions of the spatial stationarity of effects and the spatial independence of interventions (see Bergquist et al. 2023). Studies do broadly account for regional differences using regional fixed or random effect, but these procedures essentially absorb regional variation to stabilize estimates (see Vivalt 2015, 2020). However, it is that variation in impacts that needs to be connected to location specific factors in light of underlying theory to make effectiveness predictions. Capturing that variation and evaluating interventions on the timelines required to mitigate and adapt to some forms of climate change, necessitates the development of alternative evaluation systems attuned to the analysis of spatial data by a diverse community.

One path forward may be the development of long-run research programs that fuses engaged research with the quantitative spatial analysis of climate interventions. Such programs could be designed to discriminate between theories of change as evidence from new impact evaluations is produced. Nichols et al. (2021) evolving information state approach to discriminating between competing models of mallard duck populations provides one useful example of how such a continually updating system of evidence accumulation could be designed for specific climate strategies. The fundamental foundation of the Nichols et al approach is an inductive Bayesian model, which is also the foundation of the form of automated discovery and explanation proposed for geography by Gahegan (2020). GIScientists can build on these ideas, develop similar systems, and incorporate them into long-run research programs designed to progressively gather and weigh evidence about the effectiveness of climate interventions and the prospect of using those interventions in new locations. Crucially, those systems could also bring community knowledge into the evaluation and decision process through intervention design and evidence review, but also quantitatively by informing priors used in formal analyses. Precedent exists for funding such programs in the form of the National Science Foundation Long-term Ecological Research Program, while the agency's recent focus on convergent and engaged science suggest a broad recognition of the need for such work.

References

Bergquist, M., Thiel, M., Goldberg, M. H., & van der Linden, S. (2023). Field interventions for climate change mitigation behaviors: A second-order meta-analysis. Proceedings of the National Academy of Sciences, 120(13), e2214851120.

Bray, D., & von Storch, H. (1999). Climate science: An empirical example of postnormal science. Bulletin of the American Meteorological Society, 80(3), 439-456.

Cartwright, N. (2007). Hunting causes and using them: Approaches in philosophy and economics. Cambridge University Press: New York.

Cartwright, N., & Hardie, J. (2012). Evidence-based policy: A practical guide to doing it better. Oxford University Press.

Federal Geographic Data Committee (2022). FDGC Geospatial Standards. (2022, November 8). https://www.fgdc.gov/standards/list

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. Futures, 25(7), 739-755.

Gahegan, M. (2020). Fourth paradigm GIScience? Prospects for automated discovery and explanation from data. International journal of geographical information science, 34(1), 1-21.

Grace, K. (2017). Considering climate in studies of fertility and reproductive health in poor countries.

Nature climate change, 7(7), 479-485.

Harvey, D. (1969). Explanation in Geography. Edward Arnold.

Krauss, W., Schäfer, M. S., & von Storch, H. (2012). Post-normal climate science. Nature and Culture, 7(2), 121-132.

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., ... & Park, Y. (2023). IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for

Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.

Mark, D. M. (2003). Geographic information science: Defining the field. Foundations of geographic information science, 1, 3-18.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. GeoJournal, 80, 449-461.

Nichols, J. D., Oli, M. K., Kendall, W. L., & Boomer, G. S. (2021). A better approach for dealing with reproducibility and replicability in science. Proceedings of the National Academy of Sciences, 118(7), e2100769118.

Open Geospatial Consortium (2023). OGC Standards. (2023, June 30). https://www.ogc.org/standards/

Turner, B. L. (1989). The specialist–Synthesis approach to the revival of geography: The case of cultural ecology. Annals of the Association of American Geographers, 79(1), 88-100.

Turner, B. L. (2002). Contested identities: Human-environment geography and disciplinary implications in a restructuring academy. Annals of the Association of American Geographers, 92(1), 52-74.

Vivalt, E. (2015). Heterogeneous treatment effects in impact evaluation. American Economic Review, 105(5), 467-470.

Vivalt, E. (2020). How much can we generalize from impact evaluations?. Journal of the European Economic Association, 18(6), 3045-3089.

Zhining Gu, PhD student, School of Geographical Sciences & Urban Planning, Arizona State University, Tempe, AZ 85287.

Wenwen Li, Professor, School of Geographical Sciences & Urban Planning, Arizona State University, Tempe, AZ 85287.

Bing Zhou, PhD student, Department of Geography, Texas A&M University, College Station, TX 77843.

Yikang Wang, PhD student, Centre of Advanced Spatial Analysis, University College London, London, UK.

Yanbing Chen, PhD student, Department of Geography, University of Wisconsin – Madison, Madison, WI 53706.

Shan Ye, Assistant Professor, School of Information Engineering, China University of Geosciences, Beijing, China.

Kejin Wang, PhD student, Department of Environmental Sciences, Louisiana State University, Baton Rouge, LA 70808.

Hongkai Gu, PhD student, Department of Geography, Ghent University, Ghent, Belgium.

Yuhao Kang, Assistant Professor, GISense Lab, Department of Geography, University of South Carolina, Callcott, Columbia, SC 29208.

Peter Kedron, Associate Professor, Department of Geography, University of California Santa Barbara, Santa Barbara, CA, USA