Supplementary Information for: A Scoping Review of Earth Observation and Machine Learning for Causal Inference: Implications for the Geography of Poverty *

Kazuki Sakamoto † Connor T. Jerzak ‡ Adel Daoud §

September 25, 2024

Abstract

Earth observation (EO) data such as satellite imagery can have far-reaching impacts on our understanding of the geography of poverty, especially when coupled with machine learning (ML) and computer vision. Early research in computer vision used predictive models to estimate living conditions, especially in contexts where data availability on poverty was scarce. Recent work has progressed beyond using EO data to predict such outcomes but also to conduct causal inference. However, how such EO-ML models are used for causality remains incompletely mapped. To address this gap, we conduct a scoping review where we first document the growth of interest in using satellite images together with EO data in causal analysis. We then trace the methodological relationship between spatial statistics and ML methods before discussing five ways in which EO data has been used in scientific workflows—(1) outcome imputation for downstream causal analysis, (2) EO image deconfounding, (3) EO-based treatment effect heterogeneity, (4) EO-based transportability analysis, and (5) image-informed causal discovery. We consolidate these observations by providing a detailed workflow for how researchers can incorporate EO data in causal analysis going forward—from data requirement to computer vision model to evaluation metric. While our discussion focuses on health and living conditions outcomes, our workflow applies to other measures of sustainable development where EO data are informative.

Keywords: Geography; Poverty; Spatial statistics; Machine learning; Causal inference; Earth observation; Remote sensing

*We thank Ola Hall and Ibrahim Wahab for helpful comments. Thanks also go to the members of the AI & Global Development Lab for continuing discussions and inspiration. See `GitHub.com/AIandGlobalDevelopmentLab/eo-poverty-review` for a full listing of papers found in the systematic literature review.

†PhD Candidate, Institute for Analytical Sociology, Linköping University. Email: kazuki.sakamoto@liu.se; ORCID: 0000-0003-0129-3338.

‡Assistant Professor, Department of Government, University of Texas at Austin. Email: connor.jerzak@austin.utexas.edu; URL: ConnorJerzak.com; ORCID: 0000-0003-1914-8905.

§Senior Associate Professor, Institute for Analytical Sociology, Linköping University; Email: adel.daoud@liu.se; URL: AdelDaoud.se; ORCID: 0000-0001-7478-8345.

1 A Quantitative Literature Survey, Details

The search terms used were: '('eo' OR 'earth observation' OR 'satellite image*' OR 'remote sensing' OR 'earth science' OR 'earth's system' OR 'environmental eco*' OR 'environmental science*' OR 'geosci*') AND ('machine learning' OR 'ml' OR 'ai' OR 'deep learning') AND ('causal inf*' OR 'causal relation*' OR 'causality' OR 'causal learning' OR 'causal effect*' OR 'policy eval*' OR 'causal impact' OR 'causal discovery' OR 'treatment effect*' OR 'causal representation')". We limited the scope of the survey to between 2011 and 2024 (as few papers appear before 2011). This resulted in 101 papers and 41 preprints.

We analyzed the SCOPUS articles and pre-prints in the following way. First, bigrams from the abstracts were analyzed; Table **??** shows bigrams that appeared over 12 times. Beyond the expected results, some of the notable bigrams are time series, which occurred 38 times, earth system/ sciences 28 and 13, and causal discovery at 22 times. There are some trends in the type of data being processed, like time series or remotely sensed variety, the domains being studied, like earth sciences, and the types of methods or concepts, like causal discovery and Granger causality, that repeatedly show up in the nascent literature.

	Abstract Bigram	**Count**
1	machine learning	99
2	remote sensing	65
3	deep learning	60
4	time series	38
5	causal inference	34
6	earth system	28
7	causal relationships	24
8	data driven	24
9	artificial intelligence	22
10	causal discovery	22
11	neural network	21
12	causal relationship	15
13	granger causality	15
14	real world	15
15	satellite imagery	14
16	earth sciences	13
17	treatment effect	12

Table 1: *Top bigrams among papers found in the literature search.*

We then evaluate where papers are being published, Table **??** shows the various journal themes, with remote sensing journals having the greatest number followed by earth and environment, and then by AI/ML/CV publications. This goes to show that technologically minded journals have been utilized, followed by earth and environmental science-related ones. Next, the content of the paper was evaluated by themes, and surprisingly, most of the papers were grounded in earth and environmental sciences, while only 21 of the papers focused only on technical topics. For the geography of poverty, we find that there are few papers dealing with economic topics, being far eclipsed by natural science topics.

Another finding is that technical development is being outpaced by applications. The majority of papers focus on using machine learning and causal inference methods to study substantive questions in earth and environmental sciences. Comparatively, fewer papers are solely dedicated to extending

or developing novel methodologies at the intersection of these fields. This suggests there may be opportunities for more dedicated technical and methodological research to expand the toolkit for applying ML and causal inference to geospatial data.

	Journal Theme	**Count**
1	Remote sensing	18
2	Earth and environment	14
3	AI ML CV	10
4	Economics	3
5	Public health	1

Table 2: *Journal themes in quantitative review of EO-ML literature.*

In the literature found, four papers used double machine learning, five chose Granger causality, and nine used some network or graphical approach. Many of the outcomes were focused on environmental factors like vegetation, atmospheric measurements, or hydrological features. Only 12 papers were focused on socio-economic outcomes such as IWI or local development, e.g., from infrastructure projects like roads, airports, and the electric grid.

	Paper Theme	**Count**
1	Earth and environment	41
2	AI ML CV	21
3	Economics	9
4	Public health	1

Table 3: *Paper themes in quantitative review of EO-ML literature.*

Taken together, this review suggests that the intersection of causal inference, machine learning, and earth observation is a small but growing field.