AIDDATA
A Research Lab at William & Mary

WORKING PAPER 44

September 2017

A Primer on Geospatial Impact Evaluation Methods, Tools, and Applications

Ariel BenYishay
College of William & Mary and AidData

Daniel Runfola
College of William & Mary and AidData

Rachel Trichler
College of William & Mary and AidData

Carrie Dolan
College of William & Mary and AidData

Seth Goodman
College of William & Mary and AidData

Bradley Parks
College of William & Mary and AidData

Jeffery Tanner
Independent Evaluation Group, World Bank

Silke Heuser
Evaluation Department, KfW Development Bank

Geeta Batra
Independent Evaluation Office, Global Environment Facility

Anupam Anand
Independent Evaluation Office, Global Environment Facility

Abstract

The growing availability of georeferenced data on development investments and outcomes has opened up new opportunities to understand what works, what doesn't, and why at a substantially lower time and financial cost. When precisely georeferenced intervention data are fused with in-situ and remotely sensed data on outcomes like poverty, child mortality, deforestation, and governance, quasi-experimental methods of causal inference can be used to control for potential confounds and omitted variables at fine geographic levels. We introduce these geospatial impact evaluation methods, review their advantages and disadvantages, and describe their relevance and use across countries, sectors, intervention types, and development organizations.

Author Information

Ariel BenYishay
College of William & Mary and AidData
abenyishay@aiddata.wm.edu

Daniel Runfola
College of William & Mary and AidData
drunfola@aiddata.wm.edu

Rachel Trichler
College of William & Mary and AidData
rtrichler@aiddata.wm.edu

Carrie Dolan
College of William & Mary and AidData
cbdolan@wm.edu

Seth Goodman
College of William & Mary and AidData
sgoodman@aiddata.wm.edu

Bradley Parks
College of William & Mary and AidData
bparks@aiddata.wm.edu

Jeffery Tanner
Independent Evaluation Group, World Bank
jtanner@worldbank.org

Silke Heuser
Evaluation Department, KfW Development Bank
silke.heuser@kfw.de

Geeta Batra
Independent Evaluation Office, Global Environment Facility
gbatra@worldbank.org

Anupam Anand
Independent Evaluation Office, Global Environment Facility
aanand2@thegef.org

Acknowledgments

This study was made possible through generous financial support from the William and Flora Hewlett Foundation (Grant # 2015-3240) and a cooperative agreement (AID-OAA-A-12-00096) between USAID's Global Development Lab and AidData at the College of William and Mary under the Higher Education Solutions Network (HESN) Program.

Contents

Introduction

The "gold standard" in impact evaluation is the randomized control trial (RCT). An RCT uses randomization to assign exposure to a program, thereby ensuring that individual's probability of being assigned to the program is not correlated with the program's intended outcomes.[1] Individuals randomly assigned to a control group are thus statistically similar to those receiving the program and provide a strong counterfactual, allowing attribution of any differences in outcomes between the groups solely to the program's intervention.

However, RCTs are limited in their application because they often require expensive primary data collection efforts for customized samples. RCTs must also be baked into the design of programs from the outset, and implementers and evaluators must be willing to coordinate and collaborate over long periods of time. In other cases, it is impractical or unethical to randomize assignment into a program. As a result, RCTs continue to increase in use but do not yet cover most programs (Levine and Savedoff 2015; Cameron et al. 2016).

Evaluators need more tools at their disposal to rigorously measure programmatic impact when randomization is not a viable option. This article introduces geospatial impact evaluation (GIE) methods, which have opened up new opportunities to understand what works, what doesn't, and why at a substantially lower time and financial cost.

GIEs are not appropriate for all types of development interventions. However, when it is possible to (a) precisely define and measure the geographical scope of an intervention and the timing of it implementation and (b) fuse these georeferenced intervention data with in-situ and remotely sensed outcome and covariate data measured at fine geographical levels, GIEs can be a particularly useful evaluation tool. GIEs are attractive in that they can be used across a wide variety of countries, sectors, and intervention types. They can also be applied either to individual projects or project portfolios. Additionally, they offer external validity benefits because they can analyze interventions spread across entire countries (or, in some cases, over multiple countries), making it possible to draw broadly generalizable conclusions about the impacts and cost effectiveness of development interventions.

This article introduces GIE methods, reviews their advantages and disadvantages, and describes their relevance and use across countries, sectors, intervention types, and development organizations. It consists of 6 sections. Section 1 introduces identifies the conditions under which GIE methods are applicable. Second 2 outlines the benefits of GIEs. Section 3 identifies the "ingredients" that are necessary for the successful design and implementation of a GIE. Section 4 describes GIE applications in different countries, sectors, and programmatic contexts. Section 5 discusses the feasibility of using GIE tools and techniques to evaluate the impacts of different types of programs. Section 6 concludes.

---

[1] For ease of exposition, we use the term "individuals" to refer to the units of observation in an impact evaluation. These units can be individuals, communities, localities, and so forth.

Section 1: What Is Geospatial Impact Evaluation (GIE) and When Can It Be Used?

Geospatial impact evaluations (GIEs) rely on subnationally georeferenced intervention, outcome, and covariate data and quasi-experimental methods of causal inference to measure the intended (or unintended) impacts of development programs. GIEs seek to mimic the conditions of an RCT with observational data. RCTs are powerful because they create conditions under which one can reliably ascertain that individuals' participation in a program was not correlated with their outcomes. GIEs create similar conditions, but without randomly assigning individuals to treatment and control groups. The key is making comparisons across individuals who are sufficiently similar to one another and experiencing changes that are otherwise similar. The best way to make such comparisons is to identify comparison individuals who are geographically close to the program participants, but unlikely to be affected by the program's presence. Doing so requires geographically precise data on programmatic interventions and their intended (or unintended) outcomes, and as we will soon discuss, such data are rapidly expanding in number, scope, periodicity, and availability.

Rather than using randomization to identify counterfactual cases, GIEs seek to achieve a similar result through one of three methods:

(1) strategic subsampling of observational data to identify treatment and control cases that are nearly identical but for the presence or absence of the intervention (e.g. propensity score matching);

(2) comparing the pre- and post-intervention change in the outcome of interest for a treatment group relative to a control group[2] (e.g. difference-in-differences, fixed effects); or

(3) exploiting the discontinuity around a geographic cutoff that is "as-if random" (where the treated cases and control cases on either side of the cutoff are extremely similar across pretreatment covariates).[3]

GIE methods can be applied either retrospectively (for completed projects) or prospectively (for active or future projects). However, they cannot be applied to all types of development programs. The two main constraints to GIEs are the availability of data on the intended outcomes and the spatial distribution of the interventions. While outcome data are rapidly expanding in type and time periods,

[2] These approaches rely on the assumption that that the change in the control group represents the counterfactual change in the treatment group if there were no treatment.

[3] An example is when two regions on each side of a national border that were once part of the same ethnic homeland are partitioned (Michalopoulos and Papaioannou 2014). In this type of geographic regression discontinuity design, the region on the side of the border that was not subjected to the treatment might serve as the counterfactual case (if it is observationally equivalent to the "treated" region on the other side of the border across a wide array of pretreatment covariates).

they are not available retrospectively for all sectors. Secondly, GIEs are feasible for spatially differentiated interventions—those that take place in some locations but not others. A development program that provided, say, budget support or analytical and advisory support to the central government would not likely be evaluable with GIE methods. However, programs that demarcate newly protected areas, construct networks of primary health clinics, strengthen municipal governance systems, or provide agricultural extension support to farmers working on specific plots of land would likely be evaluable with GIE methods.

Section 2: The Benefits of GIEs

GIEs help to fill "the missing middle" in evaluation: they are more rigorous than performance evaluations, but significantly cheaper and faster than many randomized control trials, making it possible for a larger number of development programs to undergo impact evaluation. For programs where intervention locations have already been documented and spatially-referenced outcome and covariate data are readily accessible, a desk-based GIE can often be completed in 6-12 months at a cost of $100,000 to $150,000. By comparison, many RCTs easily take five or more years to implement and cost $500,000 to $1 million (due to the need for customized data collection in treatment and control groups at various points during the life of a program).

GIEs also have several additional advantages that make them useful. First, GIEs often make it possible to rigorously evaluate programmatic impact in cases when it is not feasible or ethical to determine which individuals or communities participate in a program through random assignment. Second, the fact that GIEs can be implemented retrospectively and remotely makes them particularly useful to evaluators working in conflict and fragile state settings. Third, GIEs can control for potential confounds and omitted variables at fine geographic levels, thereby addressing the longstanding critiques of impact evaluations that do not employ randomization methods. Of particular note is the fact that long-term data records from satellites and surveys have created new opportunities to capture pre-treatment outcome measures (e.g. land cover change, local economic development) in both treated and untreated areas.[4]

Another set of advantages that GIEs offer relate to external validity and generalizability. Whereas RCTs are often implemented in narrowly bounded settings and criticized for having weak external validity (Rodrik 2009; Ravallion 2012; Pritchett and Sandefur 2015), GIEs can involve analysis of georeferenced intervention, outcome, and covariate data for an entire country (or even multiple countries), which makes it possible to draw conclusions about the impacts and cost effectiveness of development interventions that are broadly generalizable. GIEs are also frequently based on based on panel data

[4] Accounting for pretreatment outcome levels and trends in treatment and control areas makes it easier to capture otherwise unobservable confounds that threaten causal inference (Cook, Shadish, and Wong 2008). As such, it reduces the likelihood that key confounding variables are omitted, making treated areas different from control groups even in the absence of treatment.

that cover longer periods of time than RCTs. As such, they tend to produce findings with strong external validity – in both the spatial and temporal sense.

GIE methods are also flexible tools in that that they can either be used to evaluate individual projects (e.g. Campbell et al. 2014; Buntaine et al. 2015; BenYishay et al. 2016a; Dolan et al. 2017) or project portfolios (e.g. De and Becker 2015; Buchanan et al. 2015; BenYishay et al. 2016b; Independent Evaluation Office of the Global Environment Facility 2016; Marty et al. 2017; Bunte et al. 2017). An example of the former is an evaluation of a national campaign to distribute and promote the use of long-lasting insecticide treated bednets the Democratic Republic of the Congo (DRC) (Dolan et al. 2017). This study uses data from two rounds of Demographic and Health Surveys (DHS) and exploits variation in the spatio-temporal rollout of the campaign to estimate the effect of the program on all-cause child mortality among children who were living in those provinces at the time of the campaigns. It finds that the campaign was only effective in small-geographic areas with high levels of malaria transmission. An example of the latter is a recent evaluation of a portfolio of 202 projects supported by the Global Environmental Facility (GEF) to slow, halt, or reverse land degradation. It used sub-nationally geocoded project data and satellite-based measures of land cover change, "greenness" (vegetation productivity), and land fragmentation to compare GEF project areas to an otherwise similar set of geographical areas that did not receive GEF support (Independent Evaluation Office of the Global Environment Facility 2016). This study estimated the net, attributable conservation benefits of the GEF's project portfolio and found that these interventions sequestered, on average, 43.5 tons of carbon per hectare. That amounts to roughly 108,800 tons of carbon at each GEF-funded intervention site.

Another attractive feature of GIEs is the ability to use them as the basis for calculating value-for-money (VfM) estimates -- by first translating estimates of programmatic impact into monetary values and then netting out programmatic costs. In the case of the GEF project portfolio study, it was estimated that each GEF project on average generated $7.5 million USD in carbon sequestration benefits, while the average cost of each project was $4.2 million USD.[5] The estimated rate-of-return on this investment portfolio was therefore 78.5%. In the case of the anti-malarial intervention in the DRC, Dolan et al. (2017) used their estimates of programmatic impact (child mortality reductions attributable to the anti-malarial campaign) and data on the average cost of a long-lasting insecticide treated bednet to estimate that cost of saving one child in a high malaria transmission area within the DRC is approximately $310, which represents a very cost-effective public health intervention.[6]

[5] The authors of this study used a "value transfer" approach to monetize the carbon sequestration benefits of the GEF-financed programs. They calculated the median dollar value of a ton of sequestered carbon ($12.90) from 8,093 individual valuations.

[6] This figure is based on a cost estimate of $10 per long-lasting insecticide treated bednet, a baseline monthly mortality risk of 0.2%, and a treatment effect of -0.03 percentage points (Dolan et al. 2017).

Finally, the fact that GIEs can be implemented remotely, retrospectively, and affordably opens up new opportunities to measure *long-run programmatic impacts.* RCTs often involve the collection of baseline data at the outset of a program, midline data during the implementation of the program, and endline data at program closure. However, they rarely evaluate impacts five, ten, or fifteen years after program closure due in part to the high cost of ongoing data collection for both treatment and control groups (Goldstein 2011; Bedecarrats et al. 2015: 10; Hanna et al. 2016: 82). GIEs, by contrast, often draw upon long time-series data from satellites and surveys that cover all or most locations within countries, thereby making it far easier and cheaper to track outcomes within treatment and control groups beyond the point of program closure.

This is particularly important for development programs that expect to achieve their largest impacts in the out-years (see Figure 1). Consider three brief examples. A typical theory of change for a decentralization program in a traditionally centralized governance setting would lead one to the expectation that service delivery outcomes will probably get worse before they get better (see Figure 2). Thus, from an evaluation standpoint, it would probably be most prudent to give this type of a program a relatively long "gestation period" period before collecting the final wave of endline data. Likewise, if a public health intervention is most effective at the point that "herd immunity" is achieved, the ideal time to collect endline data collection for treatment and control groups is probably not before that point. Finally, if the purpose of a large-scale investment in new road construction is to create a "growth pole" and set in motion local economic agglomeration processes, the largest benefits of this program will likely not accrue until the post-program completion period (see Figure 2).

Figure 1: A Stylized Example of Effect Sizes During the Treatment Period and the Post-Treatment Period

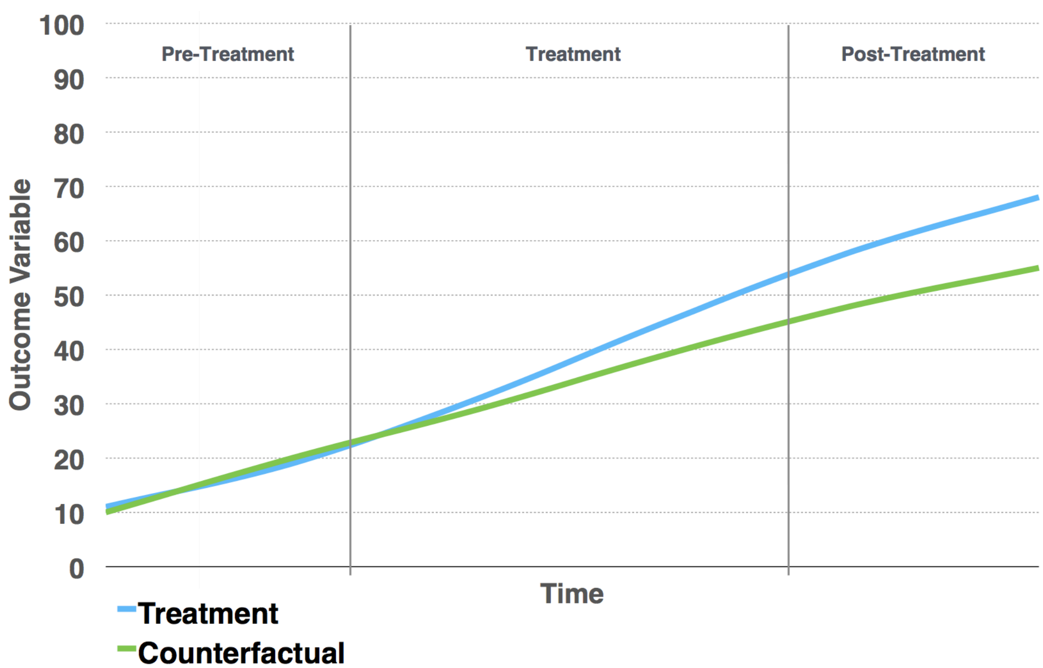
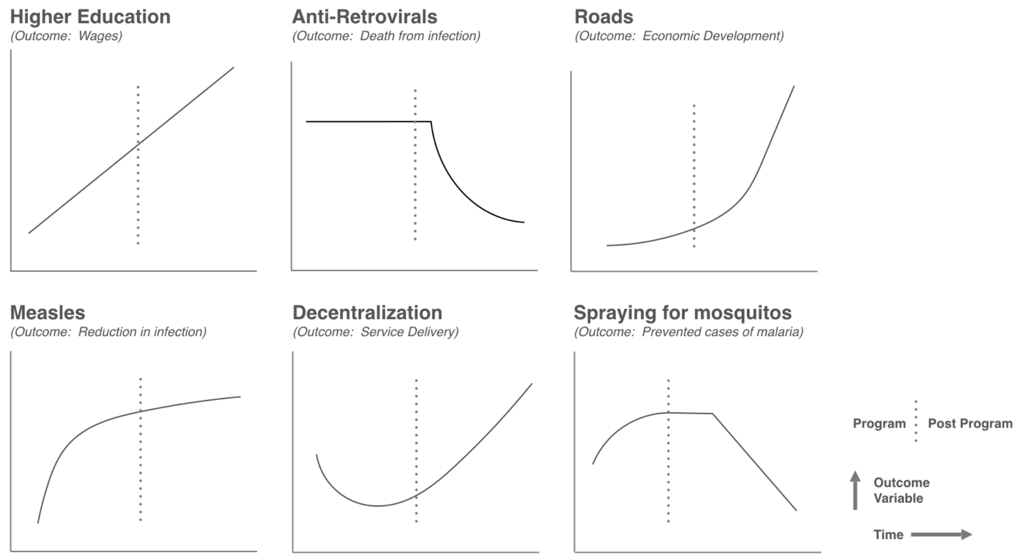


Figure 2: A Stylized Set of Programmatic Impact Trajectories[7]



Section 3: Data and Tools for Implementing GIEs

There are four key "ingredients" that are necessary for the successful design and implementation of a GIE. First, one needs precisely-defined and -measured geographic interventions and (ideally) the ability to capture variation in the geographical scope of these interventions and the timing of their implementation. Second, one needs to be able to measure the outcomes of interest and covariates at the same spatial and temporal scales. Third, one needs to be able to computationally process and join together the intervention, outcome, and covariate data at a common spatial unit of observation. Fourth, one needs econometric tools that make it possible to address the challenges of spatial uncertainty, spatial spillovers, and spatially heterogeneous effects. Here we provide a brief overview of how evaluators can bring together all of these key ingredients to successfully complete a GIE:

(1) Spatially-Explicit Intervention Data

Over the last several years, the international development community has witnessed a significant increase in the availability of geocoded intervention data (USAID 2015; AidData 2007). The World Bank now publishes the latitude and longitude coordinates of all of its investment projects. The African Development Bank, the Asian Development Bank, the United Nations Development Program, and

[7] Here we draw inspiration from Michael Woolcock's work on the functional form of development interventions (Woolcock 2009).

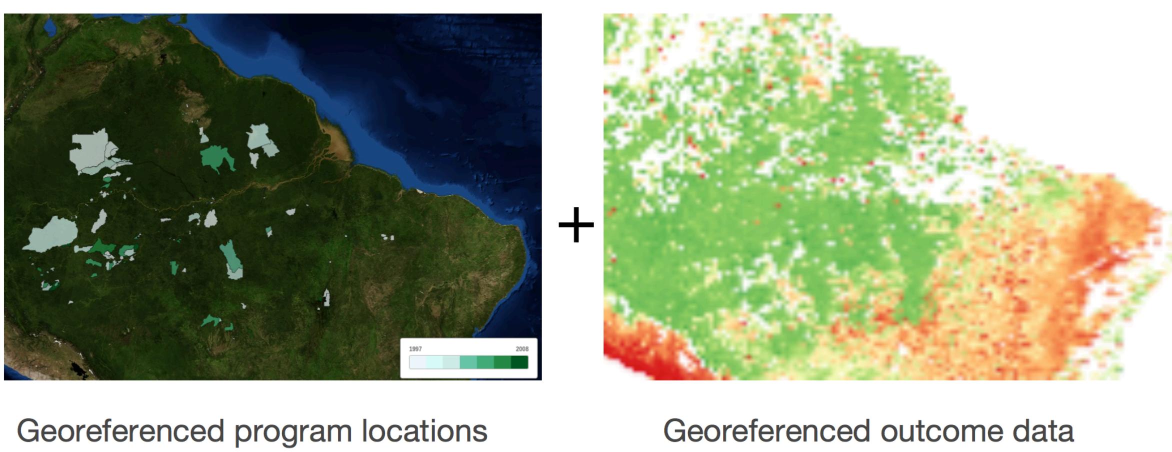
other bilateral and multilateral aid agencies have followed suit.[8] Line ministries in developing countries that are responsible for managing and coordinating incoming aid flows also increasingly publish subnationally geocoded development project data. Malawi's Ministry of Finance was the "first mover" in 2011 (World Bank 2011; Weaver et al. 2014), but nearly twenty-five finance and planning ministries in Africa, Asia, and Latin America now provide geocoded project data through their locally owned and operated aid information management systems (AidData 2017).

Even in cases when a spatially-distributed program has not georeferenced its interventions sites as points, lines, or polygons, it is often feasible to do so retrospectively and remotely. By way of illustration, in order to conduct a GIE of the Government of Liberia's spatial development corridor strategy (supported by natural resource concessions granted to foreign investors), Bunte et al. (2017) manually assembled a spatio-temporal database of natural resource sector investments by digitizing hard copies of maps from line ministries and constructing polygons based upon the field survey instructions contained in concession contracts. However, it should be noted that this additional data collection step increased the amount of money and time needed to complete the evaluation. [9]

(2) Spatially-Explicit Outcome and Covariate Data

When precisely georeferenced intervention data are fused with in-situ or remotely sensed outcome and covariate data that are measured at high levels of spatial resolution (see Figure 3 for an illustration), evaluators can use quasi-experimental techniques to control for potential confounds and omitted variables at fine geographic levels and thus address longstanding critiques of evaluations that do not employ randomization methods.

Figure 3: Joining Geocoded KFW Intervention Data With Remotely Sensed Deforestation Data in Brazil



[8] See http://maps.worldbank.org/, http://open.undp.org/, http://mapafrica.afdb.org/, and http://devgateway.github.io/asdb-gis-dashboard/.

[9] It took approximately 12 months and $100,000 to construct this dataset.

Georeferenced outcome and covariate data are expanding in number, scope, periodicity, and accessibility. A large number of Demographic and Health Surveys (DHS), Living Standards Measurement Study (LSMS) surveys, and Afrobarometer surveys have now been geocoded to the level of enumeration areas (Warren et al. 2016; Koo et al. 2016; BenYishay et al. 2017). In some cases, these micro data have also been spatially interpolated to create rasterized surfaces. A case in point is the Burke et al. (2016) child mortality decadal panel dataset constructed from DHS maternal interviews. It measures child mortality at the 10km x 10km grid cell level in the 1980s, 1990s, and 200s in in 28 Sub-Saharan African countries. In some developing countries, it is also possible to use georeferenced census data to measure socio-economic outcomes and covariates at fine geographical scales and over time (e.g. Fafchamps et al. forthcoming).

Georeferenced outcome and covariate data are also increasingly available via satellites, in situ measurement, and remotely generated event observation. The Geographically Based Economic Data (G-Econ) project provides a measure of GDP for grid cells covering the globe going back to 1990 (Nordhaus 2008). Remotely sensed nighttime light data (a proxy for subnational economic development) is available at the 1km x 1km grid cell for more than twenty-five years (Henderson et al. 2012; Bruederle and Hodler 2015; Bundervoet et al. 2015).[10] Jean et al. (2016) have developed a method for estimating consumption expenditure and asset wealth at fine geographic scales from high-resolution daytime satellite imagery. Remotely sensed data on forest cover and vegetation productivity are also accessible at fine spatial and temporal scales (Hansen et al. 2013), and it is increasingly feasible to use satellite imagery to measure agricultural productivity at the smallholder plot level (e.g. single hectares) with similar levels of accuracy to traditional field surveys, but at a fraction of the cost (Burke and Lobell 2017). Fine-scale spatial data on population, temperature, precipitation, slope, elevation, distance to roads, distance to rivers, distance to borders, distance to major population centers, natural resource deposits, and protected areas are also widely available (Goodman et al. 2016).

For evaluators who wish to understand the intended and unintended impacts of in fragile states and active conflict settings, remotely-generated observations of social and violent conflict events also provide valuable data sources. These sources include the Uppsala Conflict Data Program's Georeferenced Events Dataset (Sundberg and Melander 2013), the Armed Conflict Location and Event Database (Raleigh et al. 2010), the Integrated Crisis Early Warning System (Boschee et al. 2016), and the Social Conflict Analysis Database (Salehyan et al. 2012).

[10] Weidmann and Schutte forthcoming) demonstrate that nighttime light correlates strongly (.73) with survey-based measures of asset wealth at the local level (DHS enumeration areas with 2km-5km buffers). However, a limitation of nighttime light is that it does not do a good job of detecting small welfare changes among the extremely poor (e.g. in totally unlit areas). Jean et al. (2016) seek to overcome this challenge by using daytime satellite imagery and machine learning tools to create small-area estimates of poverty . Their method of estimating consumption expenditure and asset wealth substantially outperforms nighttime lights, with particular improvements in poorer areas.

The growing availability of *pretreatment* outcome level and trend data is particularly noteworthy, as the inability to account for pretreatment conditions in both treated and untreated areas is a key impediment to causal inference in many observation studies (Cook et al. 2008). With long term data records from satellites and georeferenced surveys, it is now increasingly possible to account for pretreatment conditions over periods as long as ten or twenty years (e.g. BenYishay 2016a; Bunte et al. 2017). By way of illustration, consider a GIE recently commissioned by the MacArthur Foundation that seeks to rigorously evaluate the conservation impacts of Chinese-funded infrastructure projects in Tanzania (BenYishay et al. 2016b). The most likely source of bias in estimating the effects of such projects on nearby forests is the siting of interventions near areas that had already experienced deforestation or near areas that would have experienced deforestation even in the absence of such projects. By way of illustration, Figure 4 shows forest loss in southeastern Tanzania in 2005 prior to the initiation of any Chinese-funded infrastructure activities. Figure 5 shows contemporaneous forest loss during a period of time in which Chinese-funded infrastructure projects were implemented. The blue dots that correspond to Chinese project sites in Figure 5 are clearly co-located with areas that experience forest loss, and in the case of the southeastern tip of Tanzania, with areas that were already experiencing deforestation *prior to Chinese-funded activities*. A cross-sectional analysis examining the change in forest cover between 2001 and 2014 would therefore detect only the positive correlation between deforestation and a grid cell's proximity to Chinese infrastructure project sites. But the evaluators responsible for this study were able to effectively expunge this source of bias by controlling for pre-treatment deforestation trends and time-varying climatic conditions at the 5km x 5km grid cell level, which led to a significantly lower (and more accurate) estimate of impact that Chinese-funded infrastructure projects had on deforestation.

Figure 4: Deforestation Pre-Trends in Southeast Tanzania (Before Program Rollout)

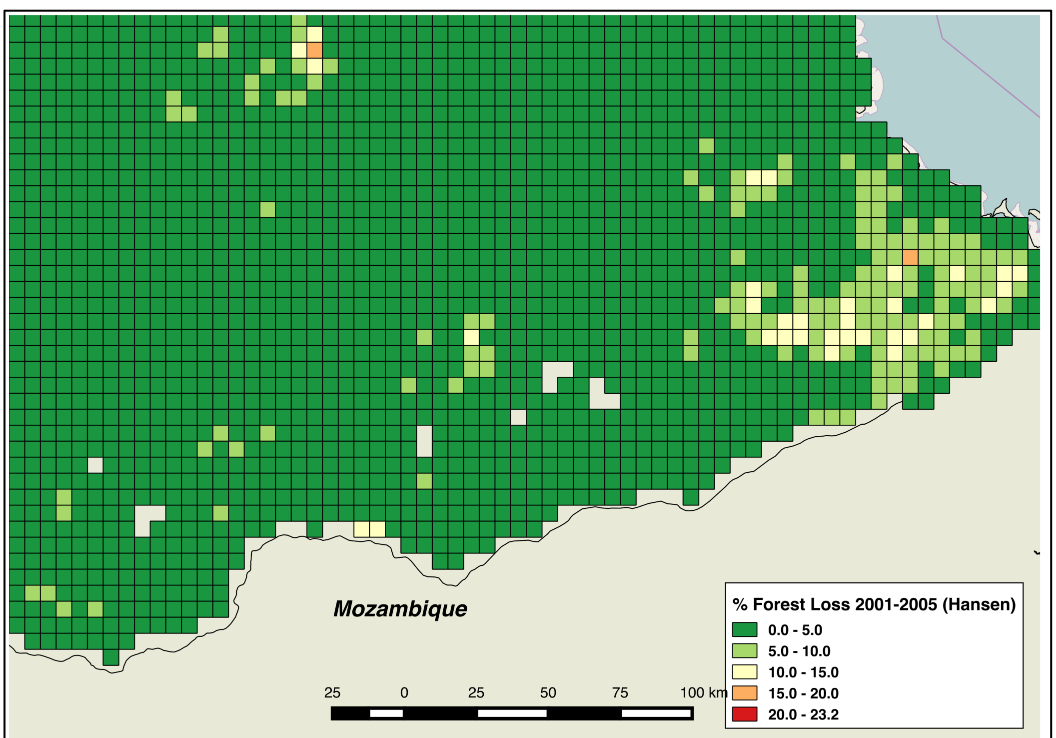
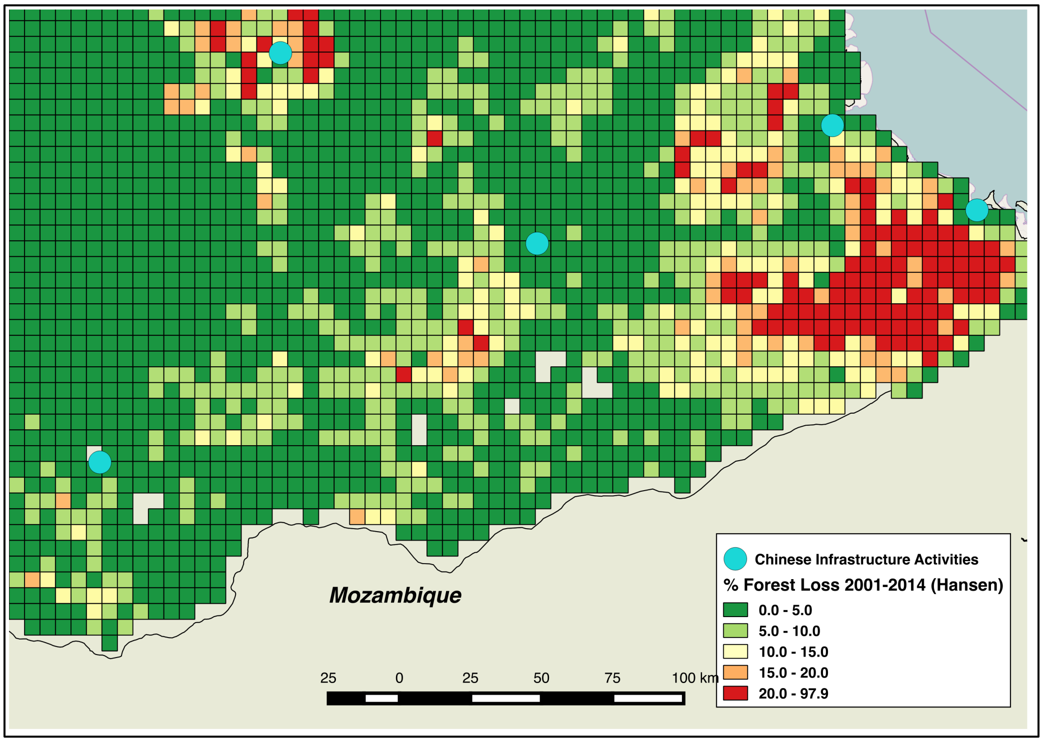


Figure 5: Deforestation in Southeast Tanzania Over Program Rollout Period



(3) The Ability to Spatially Join the Intervention, Outcome, and Covariate Data

GIEs are now significantly easier to implement because of the availability of spatial data infrastructure and increasingly affordable access to large-scale computing power. These resources make it possible to process and join together the intervention, outcome, and covariate data at a common spatial unit of observation (e.g. village, district, 1km x 1km grid cell). For example, AidData – a research lab at the College of William and Mary – has built a spatial data repository and extraction tool called geo(query) that enables evaluators to conduct GIEs at substantially lower levels of effort (Goodman et al. 2016). It allows users to obtain customized geospatial data on development interventions and outcomes without advanced GIS or computer science training. Users are able to (1) choose a country and unit of analysis (e.g. ADM1, ADM2, ADM3) and (2) select from more than two dozen spatially-referenced investment, outcome, and covariate datasets. These data are then computationally processed and fused together on the College of William and Mary's SciClone High Performance Computing Cluster at the geographical unit of analysis that the user has requested. The user is then emailed a copy of this customized dataset in a common file format (CSV, which is compatible with Excel, STATA, R, etc.). The platform, which is currently available via http://geo.aiddata.org, makes available data on:

(1) Aid: Georeferenced datasets of development projects, including donor-specific (e.g. World Bank, China), and recipient country-specific (e.g. Afghanistan, Nepal) datasets;
(2) Environment and population: Satellite- and census-based data on population density, vegetation density, protected areas, air pollution, land cover, temperature, precipitation, slope, elevation, and more;
(3) Conflict: Georeferenced conflict events from media-based and third-party sources (e.g. ACLED)
(4) Economic development and human welfare: Subnational measures of GDP, nighttime light, and child mortality; and
(5) Access to infrastructure: Spatially-referenced measures of access to markets, population centers, and roads

Going forward, AidData plans to expand the geo(query) tool to include georeferenced survey data on health and education outcomes, agricultural productivity, public service delivery, and governance (e.g. from DHS, AfroBarometer, LSMS).

(4) Econometric Tools that Account for the Unique Features of Spatial Data

Finally, for evaluators to successfully implement most GIEs, they often need econometric tools to overcome several key barriers to causal inference: spatial spillover effects, residual imprecision in the measurement of spatial data, and spatial heterogeneity in treatment effects. A set of R-based tools has recently been developed to address these common challenges. The first R package (geoSIMEX) is a method that incorporates spatial imprecision into models which seek to estimate the causal impacts of development investments (Runfola et al. 2017b). This method simulates the effect of adding measurement error to a given spatial variable. Then, once the trend in measurement error has been estimated, it back-extrapolates to conditions of no spatial measurement error. The purpose of this simulation-based method is to reduce bias in estimates of causal impact that result from use of variables that are measured with imprecision—a defining feature of many spatial data. The second R package (geoMATCH) helps researchers overcome another impediment to causal inference: geographic treatment spillover to (nearby) control units (Ho et al. 2007; Runfola et al. 2016b).[11] An example of this challenge is when a clinic may not only improve health outcomes in the geographic neighborhood where it is located, but also in nearby neighborhoods. Failing to adjust for these types of treatment spillovers can result in erroneous estimates of causal impact. Third and finally, in order to account for spatial heterogeneity in treatment effects, Zhao et al. (2016) and Runfola et al. (2017a) have developed R-based classification and regression tree routines to isolate treatment effects within sets of similar units (by classifying units of observation into clusters that are similar along covariate axes).

[11] See https://github.com/itpir/geoMatch/blob/master/README.md and http://geo.aiddata.org/docs/geoMatch.pdf

Section 4: Recent GIE Applications

In order to illustrate how GIEs are designed and implemented in practice, we now briefly summarize three recent applications.

Application 1: The PPTAL Project in Brazil

Between 1995 and 2008, the World Bank and KFW funded a Demarcation of Indian Territories project (PPTAL) in the Brazil Amazon. One of its core objectives was to reduce the risk of deforestation by physically demarcating and legally protecting 38 million hectares of indigenous lands. Seven years after project closure, BenYishay et al. (2016) conducted a GIE of the intervention. The study was possible in spite of the fact that the project was not designed or implemented with a rigorous impact evaluation in mind. In order to generate the georeferenced intervention data needed for a rigorous evaluation of the PPTAL project, the evaluators used archival records from PPTAL's implementing agency (the National Indian Foundation, or FUNAI) to measure the spatial and temporal rollout of PPTAL-funded interventions across 151 communities over 14 years. KFW also provided administrative data and project documentation on the criteria for treatment assignment. The evaluators then merged these georeferenced intervention data with 30 years of high-resolution satellite imagery on land cover outcomes and a battery of spatial covariates, including population density, distance to roads and rivers, slope, elevation, precipitation, temperature, and pre-treatment outcome levels and trends.[12] They subsequently employed quasi-experimental matching and panel methods to estimate program impacts on deforestation at the community- and grid cell-levels. Using propensity score matching and fixed effect techniques, the evaluators found little evidence of conservation effects. However, in a follow-on GIE, they set out to determine whether the demarcation reduced the incidence of land-related conflicts. With data on the annual incidence of land-related conflict in each community from yearly reports published by the Indigenous Missionary Council (CIMI) from 2003-2014 which recorded incidents of violence against indigenous peoples, they were able to estimate programmatic impacts on land conflict using panel model linear regressions with community and year fixed effects, which made it possible to control for confounds that are specific to each community and those that may have affected all communities simultaneously. They found evidence that the demarcation of indigenous lands did in fact reduce the incidence of land conflict among those communities supported early on in the PPTAL project. They also uncovered evidence that the effects of PPTAL accumulated over time, with growing protection against conflict over the years following demarcation (AidData and KFW 2016).

[12] The outcome variable in this study is the Normalized Difference Vegetation Index (NDVI). 283 billion pixels of data from AVHRR and MODIS satellites were processed on the College of William and Mary's SciClone High Performance Cluster computing cluster to create yearly aggregate summaries for each of the communities in the study (BenYishay et al. 2016).

Application 2: The UN's PBF in Burundi

From 2007 to 2013, the UN Peacebuilding Fund (PBF) supported a set of interventions in Burundi that sought to increase social cohesion within communities that hosted returning ex-combatants and internally displaced persons. Campbell et al. (2014) were later commissioned by the UN Peacebuilding Support Office (PBSO) and the PBF Joint Steering Committee (JSC) in Burundi to complete a GIE of this $44 million project portfolio. To do so, they collected household survey data from randomly sampled collines with and without PBF involvement.[13] They also used georeferenced covariate data from Burundi's national statistical office (ISTEEBU) and various line ministries. A matching algorithm was then used to identify otherwise similar locations with and without PBF activities. The study found that the PBF improved inter-group social cohesion among returning ex-combatants, IDPs, and their host communities in Bujumbura Rural, Bubanza, and Cibitoke. It also revealed heterogeneous treatment effects: otherwise similar projects worked in some locations but not in others because of differences in the ways that they were implemented (Campbell et al. forthcoming).

Application 3: A Portfolio of GEF-funded Land Degradation Projects

As the financial mechanism for the United Nations Convention to Combat Desertification (UNCCD), the GEF has supported more than six hundred projects over the past fifteen years (worth more than $3 billion) to combat land degradation worldwide.[14] In 2016, a GIE of 202 completed and ongoing GEF projects was conducted. It involved four primary steps: (1) geocoding the precise locations of GEF-financed interventions; (2) integrating remotely sensed outcome data (forest cover, forest fragmentation and vegetation productivity) with the georeferenced intervention data and a battery of spatial covariates (e.g. nighttime light, population, proximity to roads and rivers); (c) implementing a propensity score matching approach to identify locational pairs that were close-to-identical across a broad array of observable characteristics, but that differed according to whether or not they were exposed to a GEF project; and (d) employing a novel Causal Trees method to identify heterogeneous treatment effects (GEF-IEO 2016). The study found that, overall, GEF land degradation projects slowed forest loss, increased vegetation productivity, and reduced forest fragmentation within 25-kilometer catchment areas. On average, GEF projects sequestered 43.5 tons of carbon per hectare. The study also demonstrated that the initial state of the local environment is a key determinant of programmatic impact: GEF-financed interventions tend to have larger impacts in areas with poor, initial environmental conditions. Another important finding was the identification of an inflection point at which treatment effects tend to increase (4.5 to 5.5 years from project inception). The GEF also learned that its land degradation projects near urban areas tend to produce relatively small impacts

[13] Collines are the smallest administrative subdivisions in Burundi. There are more than 2500 of them and they are nested within communes and provinces.

[14] The GEF is also a financial mechanism that supports several important international environmental conventions, including the Convention on Biodiversity (CBD) and the United Nations Framework Convention on Climate Change (UNFCCC).

and provide relatively poor value-for-money (GEF-IEO 2017). Going forward, this body of evidence on the spatially and temporally heterogeneous treatment effects of past projects can inform the design and placement of future GEF projects.

Section 5: Overcoming Challenges with GIEs

The past decade has witnessed tremendous growth in the availability of subnational data (e.g. Tollefsen et al. 2012; Goodman et al. 2016), but limitations still exist in obtaining and aligning program, outcome, and covariate data at the same spatial and temporal scales – critical first steps in a GIE. Obtaining program data can be particularly challenging. Many development organizations and implementing partners still do not routinely map their intervention sites. Even when they do, information on the timing of implementation at each intervention site may be unavailable, and an overall program start and end date is rarely sufficient for this type of evaluation. The challenges of maintaining institutional knowledge and internal software systems pose additional risks to the longevity of program location and timing data, even if it once was known. Furthermore, desk-based GIEs that rely on existing spatial data can create a disconnect between implementers and evaluators – those with administrative program data may lack awareness of the opportunity to conduct a GIE, and evaluators may lack access to the spatial data collected by an implementing partner or country office and needed to identify good candidate projects.

Once georeferenced program data are in hand, the process of identifying and merging outcome and covariate data can present its own challenges. Satellites change and are replaced over time, which can impact the resolution of the data or time of day for data collection and lead to discontinuities over time (e.g. Li et al. forthcoming). These discontinuities are particularly problematic when they occur during the period of project implementation and undermine the availability of consistent pre- and post-treatment data. Geo-referenced surveys – another source of outcome data used in GIEs – may not be representative at high level of spatial resolution or their coverage may not be as comprehensive as needed (e.g. some surveys are implemented in only a few regions of a country, or only one country instead of many). Furthermore, many surveys are conducted only once, and those that are repeated are rarely done so annually, which can leave an evaluator without pre- or post-treatment data. In cases where all desired variables do exist, measurement at inconsistent spatial units of observation often requires additional time and careful decision-making to merge all variables into a single dataset.

The specific characteristics that make a project a good candidate for a GIE do bias the sectors that lend themselves to this type of evaluation in two ways. First, some sectors more easily meet the requirements of spatial precision and distribution – e.g. health, education, environment, agriculture, and infrastructure projects tend to occur in specific locations rather than diffusely throughout a larger geographic area in the way governance or capacity building projects often do. Second, project outcomes in certain sectors are more likely to be measurable through readily available spatial data. Remotely-sensed data are generally more relevant to projects with environmental or land use objectives, but less relevant for governance, education, social development, or poverty reduction projects – the main exception being nighttime lights as a proxy for economic development, though recent advances in the use of daytime imagery to capture local consumption and asset wealth outcomes are very promising (Jean et al. 2016).

Survey data help to fill sectoral gaps, but do not provide the benefits of accessibility or temporal and spatial coverage that satellite data do. Survey data also suffer from additional biases in geographic coverage related to ease of data collection (due to logistical and capacity constraints), and conflict areas are particularly susceptible to poor coverage. Evaluators may also find it harder to obtain survey data on particularly sensitive outcomes at high levels of spatial resolution.[15]

However, it should be noted that a bias toward environment, agriculture, and infrastructure projects may actually serve to broaden the impact evaluation evidence base. Cameron et al. (2016) note that 65% of IEs fall in the health, population, and nutrition sectors. GIE methods may therefore provide an opportunity to reduce this sectoral skew. Good candidate projects may be less obvious in some sectors than others, but there is still significant scope for the use of GIE methods in more sectors (as demonstrated by the examples provided in this paper).

Finally, in cases when existing geospatial program, outcome and covariate data are insufficient, it is still possible to design custom data collection for a prospective GIE, but this can increase the cost considerably. For programs where intervention locations have already been documented and spatially-referenced outcome and covariate data are readily accessible, a desk-based GIE can often be completed in 6-12 months at a cost of $100,000 to $150,000. The initial time required to examine data and determine the feasibility of a GIE is non-trivial, and starting the process with digitized program location and timing data will place a GIE on the lower end of the time and financial cost scales. GIEs that require custom data collection can raise time and financial costs closer to that of a traditional RCT ($500,000 and higher; 5 years or longer).

Section 6: Concluding Remarks

In this article, we have argued that GIE methods are useful for evaluating development programs that are geographically distributed, but they remain underutilized.

The primary benefits of GIEs are that (a) they are cheaper and faster to implement than RCTs (because they often leverage existing data rather than custom baseline and endline surveys), (b) they often produce results with strong external validity in both the spatial and temporal sense, (c) they can be conducted remotely and retrospectively across a wide variety of sectors (health, education, land rights,

[15] The additional challenges of collecting and using survey data from developing countries are well-documented (Hill 2004; Baird and Özler 2012; Das et al. 2012; Deininger et al. 2012; McKenzie and Rosenzweig 2012), and they point to the value of satellite-based measures despite their limited topical relevance.

economic development, conservation, governance), (d) they can be used to either evaluate individual projects or broader project portfolios, and (e) they make it easier to evaluate long-run impacts.

But GIE methods also have limitations. They only apply to programs that are spatially differentiated, and even among programs that meet this initial condition it is not always possible to precisely define and measure the geographical scope of an intervention and the timing of it implementation. In other cases, it is not possible to measure potential crucial confounds at the same level of spatial and temporal resolution that interventions and outcomes are observed.

At the same time, it is becoming significantly easier and cheaper to conduct GIEs because of the growing accessibility of remote sensing technologies, geo-referenced survey and census data, access to large-scale computing power, and new econometric tools. As such, GIEs are increasingly finding application within development agencies, such as USAID, the World Bank, the Global Environment Facility, the German Development Bank, and the Millennium Challenge Corporation. These organizations are using GIE methods in diverse ways – for example, to estimate the economic development impacts of local infrastructure investments, the impacts of bednet distribution programs on child survival, the state legitimacy impacts of municipal governance interventions, and the carbon sequestration impacts of biodiversity programs. GIEs thus appear to be gradually narrowing the "missing middle" in development program evaluation.

References

Addison, D., and B. Stewart. 2015. Nighttime Lights Revisited: The Use of Nighttime Lights Data as a Proxy
AidData. 2017. Subnational, Geospatial Research Datasets. Accessible at http://aiddata.org/subnational-geospatial-research-datasets

AidData and KFW. 2017. Evaluation Report: Geospatial Impact Evaluation of the Demarcation of Indian Territories Project (PPTAL) on Conflict and Deforestation. November 2016.

Baird, S. and B. Özler, B. 2012. Examining the reliability of self-reported data on school participation. *Journal of Development Economics* 98: 89–93.

BenYishay, Ariel, Silke Heuser, Daniel Runfola and Rachel Trichler. 2016a. Indigenous Land Rights and Deforestation: Evidence from the Brazilian Amazon. AidData Working Paper #22. Williamsburg, VA: AidData.

BenYishay, Ariel, Bradley Parks, Daniel Runfola, Rachel Trichler. 2016b. Forest Cover Impacts of Chinese Development Projects in Ecologically Sensitive Areas. AidData Working Paper #32. Williamsburg, VA: AidData.

BenYishay, A., Rotberg, R., Wells, J., Lv, Z., Goodman, S., Kovacevic, L., Runfola, D. 2017. Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality. Williamsburg, VA: AidData. Available online at http://geo.aiddata.org.

Boschee, Elizabeth, Lautenschlager, Jennifer, O'Brien, Sean, Shellman, Steve, Starz, James, Ward, Michael. 2016. ICEWS Coded Event Data. Harvard Dataverse, V17. Accessed at https://dataverse.harvard.edu/dataverse/icews

Bruederle, Anna and Roland Hodler. 2015. Nighttime light intensity: a proxy for economic and social development at the local level? Presented at DEVAL 'Evidence on a Silver Platter' Conference. November 2015. Berlin, Germany.

Buchanan, Graeme M., Bradley C. Parks, Paul F. Donald, Brian F. O'Donnell, Daniel Runfola, John P. Swaddle, Lukasz Tracewski, and Stuart H.M. Butchart. 2016. The Impacts of World Bank Development on Sites of High Biodiversity Importance. AidData Working Paper #20. Williamsburg, VA: AidData.

Bundervoet, T., Maiyo, L, Sanghi, A. 2015. Brights Lights, Big Cities: Measuring National and Subnational Economic Growth in Africa from Outer Space, with an Application to Kenya and Rwanda. World Bank Policy Research Working Paper, World Bank, Washington, DC.

Buntaine, M. S. Hamilton, and M. Millones. 2015. Titling Community Land to Prevent Deforestation: An Evaluation of a Best-Case Program in Morona-Santiago, Ecuador. *Global Environmental Change* 33: 32-43.

Bunte, Jonas B., Harsh Desai, Kanio Gbala, Brad Parks, Daniel Miller Runfola. 2017. Natural Resource Sector FDI and Growth in Post-Conflict Settings: Subnational Evidence from Liberia. AidData Working Paper #34. Williamsburg, VA: AidData.

Burke, Marshall, Sam Heft-Neal, and Eran Bendavid. 2016. Understanding variation in child mortality across Sub-Saharan Africa: A spatial analysis. *The Lancet Global Health* 4 (12): e936-e945.

Burke, Marshall, and David Lobell. 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences* P114 (9): 2189-2194.

Cameron, Drew B., Anjini Mishra, and Annette N. Brown. 2016. The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness* 8(1): 1-21.

Campbell, S., Dexter, T., Findley, M., Hoftmann, S., Marineau, J., and Walker, D. 2014. Independent External Evaluation, UN Peacebuilding Fund Project Portfolio in Burundi. The Graduate Institute of Geneva Centre on Conflict, Development, and Peacebuilding, http://www.unpbf.org/news/pbf-burundi-evaluation-2014/

Campbell, Susanna, Michael Findley and Danny Walker. Forthcoming. The Micro-Dynamics of Peacebuilding: A Quasi-Experimental Spatial Impact Evaluation in Burundi. Working Paper.

Cook, Thomas D, William R Shadish, and Vivian C Wong. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 27 (4): 724–750.

De, R., and C. Becker. 2015. The Foreign Aid Effectiveness Debate: Evidence from Malawi. AidData Working Paper #6. Williamsburg, VA: AidData

Das, J., Hammer, J., Sánchez-Paramo, C., 2012. The impact of recall periods on reported morbidity and health seeking behavior. *Journal of Development Economics* 98: 76–88.

Deininger, K., Carletto, C., Savastano, S., Muwonge, J., 2012. Can diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics* 98: 42–50.

Dolan, Carrie, Gordon McCord, Karen Grepin, Ariel BenYishay, and Jeff Tanner. 2017. The Impact of an Insecticide Treated Bednet Campaign on All-Cause Child Mortality: A Geospatial Impact Evaluation from the Democratic Republic of Congo. Working Paper.

Fafchamps, Marcel, Michael Koelle and Forhad Shilpi. Forthcoming. Gold Mining and Proto-Urbanization: Recent Evidence from Ghana. *Journal of Economic Geography*.

Florent Bedecarrats, Isabelle Guerin, François Roubaud. 2015. The Gold Standard for Randomized Evaluations: From Discussion of Method to Political Economy. UMR DIAL Working Paper. Paris: University of Paris-Dauphine.

Hansen, M.C., P.V. Potapov, R. Moore, M. Hancher, S.A. Turubanova, A. Tyukavina, D. Thau, S.V. Stehman, S.J. Goetz, T.R.Loveland, A. Kommareddy, A. Egorov, L. Chini, C.O. Justice, and J.R.G. Townshend. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342 (6160): 850–853

Goldstein, Markus. 2011. In the long run… Development Impact: News, views, methods, and insights from the world of impact evaluation. 06 May 2011. Accessed at http://blogs.worldbank.org/impactevaluations/in-the-long-run

Goodman, Seth, Ariel BenYishay, and Daniel Runfola. 2016. Overview of the geo framework. Working Paper. http://geo.aiddata.org/docs/IntroducingtheAidDataGeoFramework.pdf

Hanna, Rema, Esther Duflo, and Michael Greenstone. 2016. Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves. *American Economic Journal: Economic Policy* 8(1): 80–114.

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. Measuring Economic Growth from Outer Space. *American Economic Review* 102 (2): 994–1028.

Hill, Z. 2004. Reducing Attrition in Panel Studies in Developing Countries. *International Journal of Epidemiology* 33: 1-6.

Ho, Daniel Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 15(3): 199-236.

Independent Evaluation Office of the Global Environment Facility (GEF-IEO). 2016. Value for Money Analysis for the Land Degradation Projects of the GEF. Washington DC: GEF. https://www.thegef.org/sites/default/files/council-meeting-documents/EN_GEF.ME_C.51.Inf_02_VFM%20Analysis%20for%20LD%20Projects%20of%20GEF.pdf

Independent Evaluation Office of the Global Environment Facility (GEF-IEO). 2017 IEO Brief: Land Degradation Focal Area Study. Washington DC: GEF. http://www.gefieo.org/sites/default/files/ieo/signposts/files/land-degradation-2017-brief.pdf

Jean, Neal, Marshall Burke, Michael Xie, Matt Davis, David Lobell, Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353 (6301): 790–794.

Koo, Jawoo, Cindy M. Cox, Melanie Bacou, Carlo Azzarri, Zhe Guo, Ulrike Wood-Sichra, Queenie Gong, and Liangzhi You. 2016. CELL5M: A geospatial database of agricultural indicators for Africa South of the Sahara. *F1000Research* 5: 2490.

Levine, Ruth and William Savedoff. 2015. Aid at the frontier: building knowledge collectively. *Journal of Development Effectiveness* 7(3): 275-289.

Marty, R., Dolan, C. B., Leu, M., & Runfola, D. 2017. Taking the health aid debate to the subnational level: the impact and allocation of foreign health aid in Malawi. *BMJ Global Health* 2(1): e000129.

McKenzie, D., Rosenzweig, M., 2012. Preface for symposium on measurement and survey design. *Journal of Development Economics* 98: 1–2.

Michalopoulos, Stelios, and Elias Papaioannou. 2014. National Institutions and Subnational Development in Africa. *Quarterly Journal of Economics* 129 (1): 151–213.

Nordhaus, Willam D. 2008: New metrics for environmental economics: Gridded economic data. *The Integrated Assessment Journal* 8 (1): 73-84.

Pritchett, Lant and Justin Sandefur. 2015. Learning from Experiments When Context Matters. *American Economic Review* 105(5): 471-75.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. Introducing ACLED-Armed Conflict Location and Event Data. *Journal of Peace Research* 47(5) 651-660.

Ravallion, Martin. 2012. Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo's *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. *Journal of Economic Literature* 50 (1): 103- 114.

Rodrik, Dani. 2009. "The new development economics: We shall experiment, but how shall we learn?" in *What works in development? Thinking big and thinking small*, edited by Easterly, W. and Cohen, J. Washington DC: Brookings Institution Press. Pp. 24-54.

Runfola, Daniel, Ariel BenYishay, and Seth Goodman. 2016b. geoMatch: An R Package for Spatial Propensity Score Matching. R package. http://geo.aiddata.org

Runfola, Daniel, Ariel BenYishay, Jeffery Tanner, Graeme Buchanan, Jyoteshwar Nagol, Matthias Leu, Seth Goodman, Rachel Trichler and Robert Marty. 2017a. A Top-Down Approach to Estimating Spatially Heterogeneous Impacts of Development Aid on Vegetative Carbon Sequestration. *Sustainability* 9 (3): 409.

Runfola, Daniel, Robert Marty, Seth Goodman, Michael LeFew and Ariel BenYishay. 2017b. geoSIMEX: A Generalized Approach To Modeling Spatial Imprecision. AidData Working Paper #38. Williamsburg, VA: AidData.

Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. Social conflict in Africa: A new database. *International Interactions* 38 (4): 503-511.

Sundberg, Ralph, and Erik Melander. 2013. Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50 (4): 523-532.

Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug. 2012. PRIO-GRID: A unified spatial data structure. *Journal of Peace Research* 49(2): 363-374.

United States Agency for International Development (USAID). 2015. Higher Education Solutions Network - Bi-Annual Report (FY2015), AidData Center for Development Policy. Washington DC. http://pdf.usaid.gov/pdf_docs/PA00KSXN.pdf

Warren, J., Perez-Heydrich, C., Burgert, C.R., and Emch, M.E. 2016. Influence of Demographic and Health Survey Point Displacements on Distance-Based Analyses. *Spatial Demography* 4: 155.

Weaver, C., Davenport, S. , Baker, J., Findley, M., Peratsakis, C., and Powell, J. 2014. Malawi's Open Aid Map. Washington, DC: The World Bank.

Weidmann, Nils B. and Sebastian Schutte. forthcoming. Using night light emissions for the prediction of local wealth. *Journal of Peace Research*.

Woolcock. Michael. 2009. Toward a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness* 1 (1): 1-14.

World Bank. 2011. Aid Partnership: World Bank and Partners Call for Open Collaborative Maps to Improve Development Aid Impact. 25 November 2011. Accessed at http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:23054525~pagePK:34370~piPK:34424~theSitePK:4607,00.html

Zhao, Jianing, Daniel M. Runfola, and Peter Kemper. 2016. Quantifying Heterogeneous Causal Treatment Effects in World Bank Development Finance Projects. Working Paper.