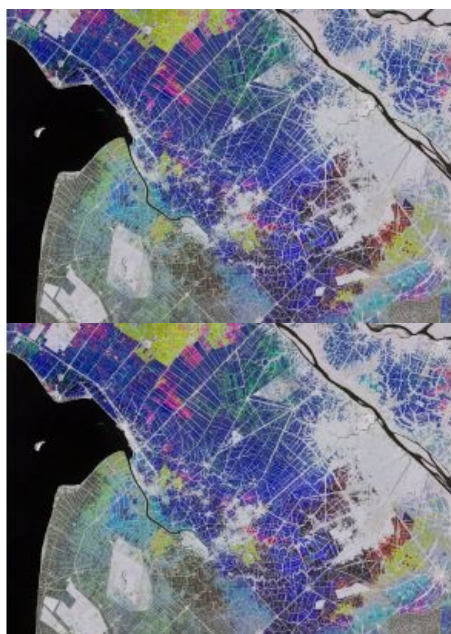


TAKING A CLOSER LOOK AT IMPACT FROM HUNDREDS OF KILOMETRES ABOVE THE EARTH

Remotely sensed data for efficient data collection



When it comes to impact evaluations, remotely sensed data can increase their timeliness, accuracy and relevance for decision-makers. 3ie and New Light Technologies are enhancing the use of geospatial analysis in IEs.

Impact evaluations (IEs) have been evolving to fill a critical gap in evidence about the effectiveness of international development programmes and interventions. Because of their ability to determine intervention effectiveness (and cost-effectiveness), the demand for and production of IEs has grown substantially in recent decades. Rigorous evaluation of development interventions and their outcomes has been a perennial challenge across multiple sectors and disciplines. Remotely sensed data allows to improve IEs in multiple ways, increasing their

timeliness, accuracy, and relevance for decision-makers. 3ie and New Light Technologies aim to enhance the generation, use and transparency of geospatial analysis in IEs.

Given the importance of international development to promote social, health, economic, and environmental well-being and equity around the world, combined with the magnitude of investments in such efforts, it is essential that we know whether or not such efforts are actually improving the targeted outcomes for the beneficiary population. Making this determination is often not straightforward, in part because it requires isolating and estimating the effect attributable to the program or intervention, as compared to what would have otherwise happened for the same population in the absence of the intervention.

The critical role of impact evaluations (IEs)

The field of impact evaluation (IE) has emerged and evolved over the past several decades to address this challenge and the associated gap in rigorous evidence on the effectiveness of development interventions. Study designs used to quantify attributable effects are typically experimental (which use random assignment to establish control groups) or quasi-experimental (which use statistical procedures to identify comparison groups to construct a valid counterfactual), often incorporating qualitative evidence as part of a mixed-methods approach.

Because of their ability to determine intervention effectiveness (and cost-effectiveness), the demand for and production of IEs has grown substantially in recent decades. Policymakers and program implementers increasingly seek evidence from rigorous IEs to guide investments toward interventions that are most likely to work, to produce the largest benefits, to reach the most people, and to do so at the lowest cost. At the same time, the limitations of IEs have also come increasingly into focus in the development community, including, for example, their often substantial time and resource costs and the challenges of accounting for important but unobserved variables and phenomena that may influence the outcomes of interest in the study population.



Satellite images are fundamental for flood detection.

The challenges and limitations of conventional IE data collection methods

Impact evaluations often rely on primary data collection or existing large-scale representative survey data to construct key outcome variables and other covariates. However, when programs are implemented at larger geographical scales than the individual or the household scale (e.g. programs implemented in villages, counties, forests, agricultural plots), conventional data collection methods may be inadequate. For example, many key outcome variables are unmeasurable by means of conventional data collection methods (e.g. small area economic activity) or are riddled with measurement errors (e.g. plot productivity).

Conventional data collection methods are also limited in their ability to measure a vast array of potentially critical control variables, such as the physical properties of areas where a program is implemented (e.g. topography, land productivity, accessibility, proximity to services, etc.), which could significantly affect the impact of the program. Furthermore, collecting several years of pre-program baseline information or conducting follow-up surveys several years after program implementation to measure long-term impacts are either prohibitively costly and/or not feasible. The high spatial and temporal resolution of satellite data allows constructing multiple comparison groups to account for spillover effects and spatial heterogeneity.

In other cases, conventional survey data collection methods, such as face-to-face interviews, prove to be challenging, for example, when it is required to reach migrant populations or populations residing in inaccessible regions. Overall, the logistical feasibility and the cost associated with the collection of conventional survey data often limits the sample size and statistical power of the analysis. When location matters for program effect, remote sensing can add enormous value to impact measurement, especially when: 1) the program placement has a spatial element (location/area/plots); 2) the outcome of interest is spatially measurable (directly or indirectly), and; 3) information on program placement and timing is available and can be clearly demarcated retrospectively.

Data collection using satellites and airborne instruments widely accessible

Until recently, the cost and availability of satellite imagery, together with the computational cost associated with data storage and analysis, have hindered the accessibility to high quality and timely satellite data for IEs. Today, there is an exponential increase in the availability of sources that provide freely accessible and reusable satellite data. Remotely sensed observations (e.g., observations collected by satellites or airborne instruments, such as drones) offer unique possibilities for IE, especially when high quality and reliable data are in short supply.

With the increasing availability, quality, granularity, and frequency of satellite data, it is now possible to collect data from almost every location on Earth. According to [UNCUSA](#), more than 2,600 satellites currently orbit Earth, with close to 40% of them collecting data specifically for Earth and space observations and for scientific applications. For example, NASA's/USGS's Landsat program has been collecting data since the 1970s, making these observations the longest continuous space-based record of Earth. The two current operating Landsat satellites (Landsat 7 and 8, which were launched in 1999 and 2013, respectively) capture, together, every location on Earth every 8 days in a spatial resolution of 30 m. Since 2014, the European Space Agency (ESA) Sentinel mission has been providing a wide range of publicly available Earth observation data, including Synthetic Aperture Radar (SAR) and electro-optical (EO) recordings. Sentinel-2, for example, provides observations of every location on Earth at a temporal frequency of up to every 5 days in a spatial resolution of 10 m. This is compared to NASA's MODIS instrument, which provides almost daily images of Earth, but in a coarser spatial resolution (down to 1km).

With terabytes of data collected by multiple sources every day, it is essential to rethink the way all this data is managed, stored and analysed. Personal computers are no longer able to process this vast amount of data. On the other hand, cloud-based computational platforms (such as Google Earth Engine, AWS, Azure, and more) now allow researchers to scale up the analysis across space and time. Parallel computing and cloud storage optimize the way the data are stored, managed and processed. With the decreasing cost of such cloud-based platforms (some of them are free for non-commercial use), it is now feasible to perform Impact evaluation in scales that were until recently impossible. These recent technology advancements provide rapid and scalable conversion of remotely sensed data into meaningful information related to economic activity, distribution of the population groups, the characteristics of land cover and land use, availability of surface water, food security, land productivity, cropping intensity and more.

Thus, combined with advancements in cloud computing and storage capabilities, this vast amount of data can be converted into meaningful information related to economic activity, distribution of the population groups, the characteristics of land cover and land use, availability of surface water, food security, land productivity, cropping intensity and more.



Detection of rice fields in a satellite image.

Remotely sensed data can help improve IEs in multiple ways

Remote sensing enables researchers to measure outcomes and construct comparison groups in ways and scales that until recently were not possible to meet the needs of policymakers. It strengthens the analysis, for example, through controlling for confounders and pre-program trends and makes the evaluations more feasible, for example, due to the reduction of data collection costs and more cost-effective retrospective and remote analysis.

1. *Measuring the unmeasurable:*

- **Measuring outcomes:** Outcomes such as economic growth, GDP, poverty or wealth at the sub-national level, infrastructure quality, population distribution, etc. are difficult to measure accurately and/or at a required temporal and spatial scale using conventional data collection methods. Remotely sensed night light data may serve as a proxy for these outcomes. For example, human-generated light at night is used as a proxy for local area economic activity.
- **Constructing the comparison group:** Remotely sensed data enables matching comparison units based on relevant pre-program characteristics at the appropriate unit level or based on spatial discontinuity. A common method to identify comparison groups is performing pipeline or sequential allocation, where untreated segments function as a comparison group until they are treated; a regression discontinuity design includes units within a specified cut-off (e.g., within a given radius around the program), thus creating

a comparison group from a pre-specified contiguous space. The fact that satellite data covers all areas and is always “on” (both temporally and spatially) also means that it is not susceptible to self-selection bias like other sources of big data (e.g., call detail records, interest-based search online searches, social media data, etc.).

- **Long term impact:** Collecting several years of pre- and post-program data through a face-to-face survey is expensive and, in many cases, infeasible. The possibility to collect pre- and post-program data, especially the follow-up data, without the need for going to the field, enables measurement of the long-term program impact, and can help analyze how the impacts evolve over time and how long they last.

2. Overcoming analytical challenges:

- **Assessing pre-program trends:** Quasi-experimental designs require pre-program similarity between the treatment and the control group, both in levels and distribution, potentially for several years before the program. Historical time series of satellite data makes it possible to evaluate parallel trend assumptions.
- **Controlling for covariates:** Failing to control for confounding factors will lead to omitted variable bias. Remote sensing can help control for local area, time-varying factors through fixed effects at the level of individual cells or pixels in aerial or satellite imagery, and for time-invariant factors, such as physical attributes, through directly measuring them.
- **Heterogeneous effects:** IEs often measure the average treatment effect for an entire treatment group rather than heterogeneous effects for sub-groups, largely due to the unavailability of data and statistical power limitations. Remote sensing allows researchers to estimate heterogeneous effects based on observable baseline conditions such as population density at the cell-level, etc. with sufficient power for sub-group analysis.
- **Robustness analyses:** Remotely sensed data can help conduct robustness analyses by allowing for the identification of multiple comparison groups that would have been expensive or infeasible through traditional data collection methods. Similarly, placebo tests can be conducted through testing the treatment effect on the treated for an arbitrary pre-program date.
- **External validity and generalizability:** Remotely sensed data are available not only for the program area, but also for the country/regional context. However, one needs to be mindful of the challenges in generalizing it beyond the country from which the training data comes from.

3. Overcoming logistical challenges:

- **Cost of data collection:** A fundamental challenge of IEs is the cost of survey data collection. For example, the average cost of a 3ie-funded multi-year, multi-round survey impact evaluation is approximately USD 400k, where the survey alone costs USD 175k. In comparison, the cost of a desk-based impact evaluation with free remotely sensed data would be around USD 150k.
- **Retrospective, desk-based evaluation:** For certain types of programs, historical time series satellite data allows retrospective assessment of interventions already implemented and in most cases the evaluation can be implemented remotely.

The key limitations of remotely sensed data

For some applications, such as counting the number of trees or detecting building footprints, there is a need for the highest possible spatial resolution; for other applications, the temporal or spectral resolution may prove to be more important (for example, a higher spectral resolution will be necessary to automatically detect types of crop fields and a high temporal resolution will be essential to monitor daily changes in agricultural land productivity). In general, there is an inherent trade-off between each of these characteristics. For example, high spatial resolution imagery will often have a lower spectral resolution, while high spatial resolution imagery will often be associated with a lower temporal resolution, the covered imagery will be smaller and the imagery will be costlier.

Nonetheless, the number and types of commercial imaging satellites are continuing to increase while small and micro satellites become significantly cheaper to build and launch. Simultaneously, the cost associated with data storage, management, and analysis is continuing to decrease. This could potentially revolutionize the field of impact evaluation, which, in many cases, requires imagery in the highest possible spatial resolution (for example, a spatial resolution of 10m will not be sufficient for counting the number of trees in a small agriculture field).



Band combination techniques can be used to highlight green vegetation in a satellite image.

Despite the increasing availability and use of satellite data, there are some important considerations that must be taken into account, including cloud coverage (which may limit the collection of remotely sensed data in the tropics), the revisit period of the satellites (which tends to be lower at the spatial resolution increases), sensor limitations (e.g. the failure of Landsat's-7 Scan Line Corrector (SLC) in 2003, which resulted in significant data gaps in the acquired scenes), and the need for robust and scientifically sound data cleaning and post-processing. Importantly, it is essential to take the necessary steps in order to be able to make assumptions and generalize the interpretations. For example, machine learning and artificial intelligence approaches are often used to convert remotely sensed data into meaningful information about the Earth. Some of these approaches rely on supervised machine learning techniques, which require reference data for training (and validation). It is important to ensure the generalization of the reference data; for example, reference data for any supervised image classification must be collected from diverse geographical regions and a wide-range of examples.

From an ethical perspective, it is important to consider risks of re-identification of study subjects and infringement of privacy, particularly

if/when using ultra-high resolution images (e.g., less than 1m). In practice, these risks can be mitigated by using areas or regions as the unit of analysis or using geomasking methods to protect privacy while maintaining spatial resolution.

Remote sensing as a complement

As implied by the above considerations, remote sensing should be approached as a complement, rather than a replacement to, conventional forms of data collection, with an emphasis on those aspects of IEs where it is most likely to add value. One initial and perhaps obvious consideration is the type of intervention and research question, as remote sensing will be much more relevant for some than others. For instance, satellite or airborne imagery of agricultural fields are likely to be more accurate, unbiased, affordable, and faster for measuring crop yields than self-reports or direct observation (e.g., crop cuts). In contrast, images from above will not tell us much about interventions to improve employee productivity and satisfaction within a workplace, nor will they help us understand (going back to agricultural productivity) why crop yields are changing and the underlying causal mechanisms.

A second important consideration is that there are multiple critical aspects of development research that are more difficult to do remotely. For example, meaningful and ongoing engagement with policymakers, implementers, beneficiaries, and other stakeholders are essential to ensure that an impact evaluation is responsive to local needs and policy questions. Despite a growing use of virtual meeting platforms, this is still difficult (and sometimes impossible) to do effectively from afar. An increasing emphasis on process evaluation and mixed methods impact evaluations similarly highlights the limitations of remote sensing, as satellites and airborne images do not provide qualitative information about the political, social, and operational context in which an intervention is conducted or how it is understood and experienced by implementers and beneficiaries.

Conclusion

Rigorous evaluation of development interventions and their outcomes has been a perennial challenge across multiple sectors and disciplines. This is due, in part, to the fact that conventional evaluation methods, such as household surveys, are typically costly, time-consuming, and often unable to capture important spatial aspects of these programmes. Remotely sensed data, such as satellite and aerial imagery, can contribute significantly to increasing the efficiency of data collection for some variables and opening up the possibility of accounting for others that were previously so onerous to collect or meaningfully synthesize that they were effectively “unmeasurable”. However, a naïve use and interpretation of these data in IEs may result in misleading conclusions. As we show in this paper, a number of technical challenges such as cloud coverage, time of data capture, data gaps due to technical glitches and data comparability must be accounted for carefully. Furthermore, the need to validate the predictions and interpretation (e.g., by means of machine learning techniques), the continued importance of meaningful stakeholder engagement, and the growing emphasis on process evaluations and mixed methods to understand implementation and context point to the complementarity between conventional surveys and remotely sensed data, rather than the latter replacing the former.

How 3ie and NLT can help customers with impact evaluation

In light of the increasing demand for geospatial analysis in impact evaluation, the rapid recent advancements in access to geospatial and remotely sensed data and the development of new methods to convert this data into information that is fundamental of IE, the International Initiative for Impact Evaluation (3ie) and [New Light Technologies](#) Inc. have partnered to enhance the generation, use, and transparency of geospatial analysis in impact evaluation, with an emphasis on informing development decision-making and strengthening research capacity in low- and middle-income countries. Together, they aim to enhance the generation, use, and transparency of geospatial analysis in impact evaluation, with an emphasis on informing development decision-making and strengthening research capacity globally, and in particular stakeholders in low- and middle-income countries.

<https://www.gim-international.com/content/article/remotely-sensed-data-for-efficient-data-collection>
